New Perspectives and Methods in Loss Reserving Using Generalized Linear Models

Jian Tao

A Thesis

for The Department of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics) at

Concordia University

Montréal, Québec, Canada

September 2014

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Jian Tao**

Entitled: **New Perspectives and Methods in Loss Reserving Using Generalized Linear Models**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Examiner
Dr. P. Gaillardetz

______________________________ Examiner
Dr. Y. P. Chaubey

______________________________ Thesis Supervisor
Dr. J. Garrido

Approved by ______________________________
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date ______________________________

ABSTRACT

New Perspectives and Methods in Loss Reserving
Using Generalized Linear Models
by Jian Tao

Loss reserving has been one of the most challenging tasks that actuaries face since the appearance of insurance contracts. The most popular statistical methods in the loss reserving literature are the Chain Ladder Method and the Bornhuetter Ferguson Method.

Recently, Generalized Linear Models (GLMs) have been used increasingly in insurance model fitting. Some aggregate loss reserving models have been developed within the framework of GLMs (especially Tweedie distributions). In this thesis we look at loss reserving from the perspective of individual risk classes. A structural loss reserving model is built which combines the exposure, the loss emergence pattern and the loss development pattern together, again within the framework of GLMs. Incurred but not reported (IBNR) losses and Reported but not settled (RBNS) losses are forecasted separately. Finally, we use out of sample tests to show that our method is superior to the traditional methods.

In the third chapter we also extend the theory of limited fluctuation credibility for GLMs to one for GLMMs. Some criteria and algorithms are given. This is a byproduct of our work but is interesting in its own sake. The asymptotic variance of the estimators is derived, both for the marginal mean and the cluster specific mean.
Keywords: GLMs, GLMMs, IBNR, RBNS, UMSEP, asymptotic variance, full credibility, loss reserving, individual risk classes.

Acknowledgments

First and foremost, I wish to express my utmost gratitude to my supervisor Dr. José Garrido for his encouragement, painstaking support and guidance throughout my study at Concordia University.

I also would like to take the opportunity to thank for Dr. Patrice Gaillardetz and Dr. Yogendra Chaubey for serving on my thesis committee.

Besides, I would like to thank all my colleagues and the assistants in the department who helped me a lot, not only on academic matters but also in life.

Last but not least, I would like to thank my parents for bringing me up in China, my aunt and her family for taking care of me in Montreal.

Contents

List of Tables

List of Figures

Glossary

Accident year (AY) the relative year to the beginning of the business or the beginning of the data available (base year) in which a claim incurred, starting from 1.

Case reserve an estimate of the amount for which a particular claim will ultimately be settled or adjudicated.

Exposure the measurable extent of risk, for instance it could refer to the number of insurance contracts in one accident year.

IBNR Incurred but not reported.

Loss reserve an estimate of the value of a claim or group of claims not yet paid.

Payment delay (PD) the relative year to the notification of a claim in which a payment was made for that claim, starting from 0.

Payment year (PY) the relative year to the occurrence in which payments were made for one claim, starting from 0.

RBNS Reported but not settled.

Reporting delay (RD) the relative year to the occurrence of a claim in which the claim was reported, starting from 0.

Settlement delay (SD) the relative year to the notification of a claim in which the claim was closed, starting from 0.

Settlement year (SY) the relative year to the occurrence in which one claim was settled, starting from 0.

Symbols

$N_{i,j}^{(k)}$ number of claims in risk class k with $AY = i$, $RD = j$, $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, m$ and $j = 0, 1, \ldots, m-1$.

$S_{i,j}$ the cumulative paid losses of accident year i and up to payment year j for the whole portfolio, $i = 1, 2, \ldots, m$ and $j = 0, 1, \ldots, m-1$.

$Z_{i,j}$ the incremental paid losses in accident year i and payment year j for the whole portfolio, $i = 1, 2, \ldots, m$ and $j = 0, 1, \ldots, m-1$.

$\mathcal{N}$ the set $\{N_{i,j}^{(k)} \mid k = 1, 2, \ldots, K,\ i = 1, 2, \ldots, m \text{ and } j = 0, 1, \ldots, m-1\}$.

$\mathcal{N}^{(k)}$ the set $\{N_{i,j}^{(k)} \mid i = 1, 2, \ldots, m \text{ and } j = 0, 1, \ldots, m-1\}$ for $k = 1, 2, \ldots, K$.

$\mathcal{N}_i$ the set $\{N_{i,j}^{(k)} \mid k = 1, 2, \ldots, K, \text{and } j = 0, 1, \ldots, m-1\}$ for $i = 1, 2, \ldots, m$.

$p_j^{(k)}$ probability for one specific claim in risk class k to be reported with $RD = j$, $k = 1, 2, \ldots, K$ and $j = 0, 1, \ldots, m-1$.

$w_i^{(k)}$ exposure for risk class k in accident year i, $k = 1, 2, \ldots, K$ and $i = 1, 2, \ldots, m$.

$\mathbf{x}_p^{(k)}$ claim payment covariates for risk class $k = 1, 2, \ldots, K$.

$\mathbf{x}_s^{(k)}$ claim severity covariates for risk class $k = 1, 2, \ldots, K$.

$\mathbf{x}_f^{(k)}$ claim frequency covariates for risk class $k = 1, 2, \ldots, K$.

$\mathbf{x}_r^{(k)}$ claim reporting delay covariates for risk class $k = 1, 2, \ldots, K$.

$\boldsymbol{\beta}_f$ regression coefficient vector for claim frequency.

$\boldsymbol{\beta}_p$ regression coefficient vector for claim payments.

$\boldsymbol{\beta}_r$ regression coefficient vector for claim reporting delay.

$\boldsymbol{\beta}_s$ regression coefficient vector for claim severity.

Chapter 1

Introduction to Loss Reserving

1.1 Introduction to Run-off Process

Figure 1.1 illustrates the run-off (development) process of a general insurance claim. A claim occurs at a certain point t_1, consequently it is reported to the insurer at t_2 and one payment, several payments or no payment may follow until the settlement of the claim at t_6.

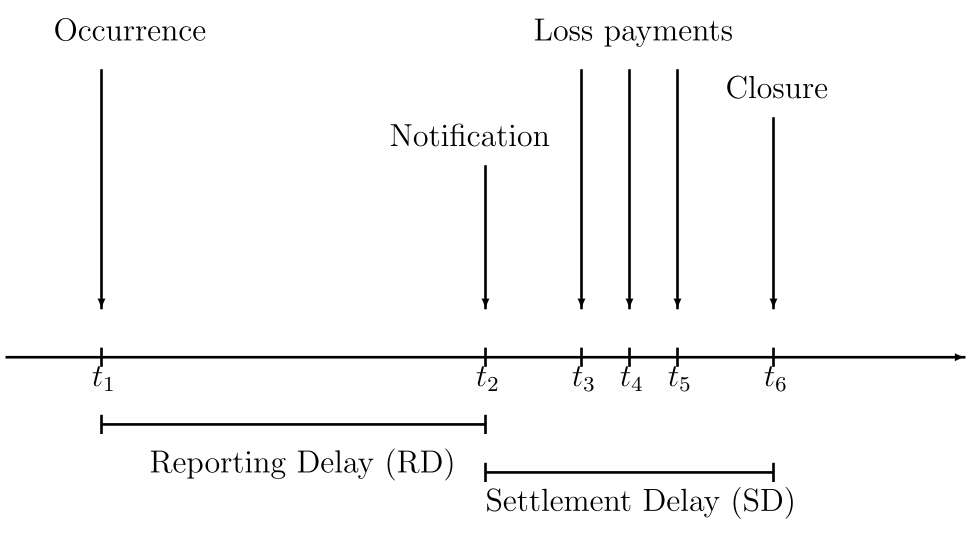

Figure 1.1: Development of a General Insurance Claim.

Most of the time, we do not know the exact time point, but only the calendar year (or month, quarter) these actions fall into. Suppose that the reference is year 2000 (base year), an accident happened in 2003, it was then reported to the insurer in 2008, one payment was made in 2008 and another payment in 2010, settling the claim in that same year. Then we know that for this specific claim that $AY = 4$, $RD = 5$, $SD = 2$, and that there are two payment delays: PD_1=0, PD_2=2.

If we focus on the claim emergence and claim reporting patterns, Table 1.1 gives us a general picture of the reporting process for all claims. The entries $N_{i,j}$ denote the number of reported claims in the portfolio that happened in accident year i and notified to the insurer with a reporting delay of j. Most papers in the literature set the upper-bounds on the accident year and the reporting delay to be equal (i.e. $I = J$), so we are more familiar with the sub-table below the dash line, which is a right-angle isosceles triangle, however Table 1.1 describes the general situation.

Accident	Reporting Delay						
Year	0	1	$\cdots$	j	$\cdots$	$J-2$	$J-1$
1	$N_{1,0}$	$N_{1,1}$	$\cdots$	$N_{1,j}$	$\cdots$	$N_{1,J-2}$	$N_{1,J-1}$
2	$N_{2,0}$	$N_{2,1}$	$\cdots$	$N_{2,j}$	$\cdots$	$N_{2,J-2}$	$N_{2,J-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I+1-J$	$N_{I+1-J,0}$	$N_{I+1-J,1}$	$\cdots$	$N_{I+1-J,j}$	$\cdots$	$N_{I+1-J,J-2}$	$N_{I+1-J,J-1}$
$I+2-J$	$N_{I+2-J,0}$	$N_{I+2-J,1}$	$\cdots$	$N_{I+2-J,j}$	$\cdots$	$N_{I+2-J,J-2}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\iddots$		
$I-j$	$N_{I-j,0}$	$N_{I-j,1}$	$\cdots$	$N_{I-j,j}$			
$\vdots$	$\vdots$	$\vdots$	$\iddots$				
$I-1$	$N_{I-1,0}$	$N_{I-1,1}$					
I	$N_{I,0}$						

Table 1.1: Aggregate Report Table

The core idea of our individual method is that in this portfolio, policyholders are classified according to their attributes (covariates, predictive variables) into different risk classes $k = 1, 2, \ldots, K$. Hence we can draw the individual risk class version of Table 1.1 by adding a superscript k to each $N_{i,j}$.

The claim payment process is a little more complicated than the reporting process (frequency), since we are not dealing with count data, but continuous data that usually exhibit a larger variance. One strategy is to first calculate the total losses associated with each claim that has currently settled (i.e. at evaluation time). Projections of total losses for future reported individual claim are based on these settled claims. Combined with the projection of future reported claim numbers, we can then give IBNR losses for each risk class. Finally adding them up, we obtain the total IBNR losses for the whole portfolio.

Some decision makers are also curious about the way future total losses are distributed to each payment year so reserves can be set year by year dynamically. Note that at any given time there might be some claims that have been reported and may have initiated some loss payments but that are not fully settled yet. That is another reason for which we want to model the claim payment pattern and picture the RBNS losses.

1.2 Literature Review

1.2.1 Chain Ladder Type Methods

1.2.1.1 Pure chain ladder method

The most widely used method for loss reserve projections is the chain ladder method, due to its simplicity and the fact that it is distribution free. Here and henceforth for the ease of exposition and without loss of generality, we set the accident year (AY) $i = 1, 2, \ldots, m$ and payment year (PY) $j = 0, 1, \ldots, m-1$. If we refer to $S_{i,j}$ as the cumulative paid losses of accident year i and up to payment year j, for the whole portfolio (see Table 1.2), then the corresponding observed incremental losses are given by

$$Z_{i,j} = \begin{cases} S_{i,0} \ , & \text{if } j = 0, \\ S_{i,j} - S_{i,j-1} \ , & \text{if } 1 \leq j \leq m-i. \end{cases} \tag{1.1}$$

The chain ladder technique estimates the corresponding development factors by

Accident	Payment Year								
Year	0	1	$\cdots$	j	$\cdots$	$m-i$	$\cdots$	$m-2$	$m-1$
1	$S_{1,0}$	$S_{1,1}$	$\cdots$	$S_{1,j}$	$\cdots$	$S_{1,m-i}$	$\cdots$	$S_{1,m-2}$	$S_{1,m-1}$
2	$S_{2,0}$	$S_{2,1}$	$\cdots$	$S_{2,j}$	$\cdots$	$S_{2,m-i}$	$\cdots$	$S_{2,m-2}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$			
i	$S_{i,0}$	$S_{i,1}$	$\cdots$	$S_{i,j}$	$\cdots$	$S_{i+1,m-i}$			
$\vdots$	$\vdots$	$\vdots$		$\vdots$					
$m-j$	$S_{m-j,0}$	$S_{m-j,1}$	$\cdots$	$S_{m-j,j}$					
$\vdots$	$\vdots$	$\vdots$							
$m-1$	$S_{m-1,0}$	$S_{m-1,1}$							
m	$S_{m,0}$								

Table 1.2: Chain Ladder Triangle

$$\hat{D}_j = \frac{\sum_{i=1}^{m-j} S_{i,j}}{\sum_{i=1}^{m-j} S_{i,j-1}}, \qquad j = 1, 2, ..., m-1 \ , \tag{1.2}$$

then the projection up to the jth payment year of the total paid losses is given by:

$$\hat{S}_{i,j} = S_{i,m-i} \cdot \hat{D}_{m+1-i} \cdot \hat{D}_{m+2-i} \cdots \hat{D}_j, \qquad i + j > m, \tag{1.3}$$

while the predictor of incremental paid losses is obtained by differences:

$$\hat{Z}_{i,j} = \begin{cases} \hat{S}_{i,m+1-i} - S_{i,m-i}, & \text{if } j = m + 1 - i, \\ \hat{S}_{i,j} - \hat{S}_{i,j-1}, & \text{otherwise.} \end{cases} \tag{1.4}$$

For a long time, no statistical model justified this method until Renshaw and Verrall (1998) found that the chain ladder projection could be interpreted as the result of a Poisson regression with categorical variables for accident years and payment years.

In their model, it is assumed that the incremental losses $Z_{i,j} \sim$ Poisson with mean $\mu_{i,j}$, independently $\forall i, j$, where $\log \mu_{i,j} = \mu + \alpha_i + \beta_j$. Here the reference parameters satisfy: $\alpha_1 = \beta_1 = 0$. Projections are done as follows:

$$\hat{Z}_{i,j} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j} \tag{1.5}$$

and

$$\hat{S}_{i,j} = S_{i,m-i} + \sum_{l=m-i+1}^{j} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_l}, \tag{1.6}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_j$ are the maximum likelihood estimators (MLE) of the parameters α_i and β_j.

At first sight it seems that explicit formulas for the predictor are impossible. However, Renshaw and Verrall (1998) found that these predictors give the same result as the chain ladder technique. The problem arises since Poisson regression is more appropriate for count data, thus it is more reasonable to model the reported or paid claim number $N_{i,j}$ rather than total losses $Z_{i,j}$ as in (1.5).

Rosenberg (1990) developed a method for modelling the claim reporting or settlement pattern. Denote $p_{(i)j} = \frac{p_j}{\sum_{k=0}^{m-i} p_k}$ the conditional probability that a claim with accident year i, known to have been reported or settled, is reported or settled with $RD = j$. Rosenberg (1990) writes the likelihood function through the multinomial distribution to get the MLEs $\hat{p}_j$:

$$L_c = \prod_{i=1}^{m} \Big\{ \frac{C_{i,m-i}}{\prod_{j=0}^{m-i} N_{i,j}!} \prod_{j=0}^{m-i} p_{(i)j}^{N_{i,j}} \Big\}, \tag{1.7}$$

where $C_{i,m-i} = \sum_{i=1}^{m-i} N_{i,j}$. Then the projection for total claim number $C_{i,m-1}$ in accident year i is given by:

$$\hat{C}_{i,m-1} = \frac{C_{i,m-i}}{1 - \sum_{j=m-i+1}^{m-1} \hat{p}_j}. \tag{1.8}$$

The projected incremental claim number $\hat{N}_{i,j}$ is given by:

$$\hat{N}_{i,j} = \hat{C}_{i,m-1} \cdot \hat{p}_j, \qquad i + j > m. \tag{1.9}$$

This approach reproduces the chain ladder method (CLM) for claim number triangle, that is to say it is equivalent if we replace the entry $S_{i,j}$ in the run-off triangle in Table 1.2 with $C_{i,j}$ and apply the chain ladder projection.

1.2.1.2 Bornhuetter Ferguson method

In the method of Bornhuetter and Ferguson (1972) (BF), it is assumed that there exist parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$ and $\gamma_0, \gamma_1, \ldots, \gamma_{m-1}$, with $\gamma_{m-1} = 1$, such that:

$$\mathbb{E}[S_{i,j}] = \alpha_i \gamma_j, \qquad \text{for all } i = 1, 2, \ldots, m \text{ and } j = 0, 1, \ldots, m-1. \tag{1.10}$$

Thus $\alpha_i = \mathbb{E}[S_{i,m}]$ represent the expectation of total losses in accident year i (row effect), and $\gamma_0, \gamma_1, \ldots, \gamma_{m-1}$ form the development pattern (column effect). This method is based on the prior estimators $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_m$ and $\hat{\gamma}_0, \hat{\gamma}_1, ..., \hat{\gamma}_{m-1}$ where $\hat{\gamma}_{m-1} = \gamma_{m-1} = 1$. These prior estimators could be obtained from internal information which is contained in the run-off triangle and external information obtained from market statistics for some other similar portfolios. In this sense, any complex projection using external information belongs to the scope of the BF method. However, in their paper Bornhuetter and Ferguson considered the aggregate cumulative run-off triangle, as in Table 1.2, and the cumulative losses projection $\hat{S}^{BF}_{i,j}$ is defined straightforwardly as:

$$\hat{S}^{BF}_{i,j} = S_{i,m-i} + (\hat{\gamma}_j - \hat{\gamma}_{m-i})\hat{\alpha}_i, \qquad \text{for all } i = 1, 2, \ldots, m \text{ and } j = 0, 1, \ldots, m-1. \tag{1.11}$$

From this formula we can see that for those accident years that convey less information, i.e. i is close to m, the prior information is dominant in the total losses projection. Prior information is especially useful when we find that the data is poor and unreliable. So the BF method solves the well known weakness of the CLM against outliers and it is more robust than the CLM which relies completely on the data contained in the run-off triangle.

1.2.1.3 Munich chain ladder

Reserves for a portfolio are often calculated on the basis of a paid losses run-off triangle for most of the methods. Sometimes we can also use case reserves to create a reported losses run-off triangle. Most decision makers choose one between paid losses and case reserves and neglect the information from the other or treat them separately. Quarg and Mack (2008) criticize this separate chain ladder (SCL) for the following reasons:

1. Most of the time the projection based on paid losses differs from the projection from reported losses. There is no strong argument to select one triangle over the other.

2. Projections based on paid losses triangles arbitrarily ignore the fact that large reported losses will lead to large paid losses in the future.

3. Projections based on reported losses use case reserves (predictions of claim amounts), not true paid losses, thus often leading to bias.

4. If we extrapolate both the paid and the reported triangles, denoted as P and I (known as incurred), use the chain ladder method and create the associated (P/I) quadrangle by dividing each term in these two quadrangles, then the weakness of SCL becomes clearly apparent. Usually an above-average or below average (P/I) will lead to an above-average or below average projection (P/I) at the end of the quadrangle. Some years the projection (P/I) at time m will

be greater than 100%, other years the ratio will be far less than 100%. Both cases contradict what is observed in practice.

Hence Quarg and Mack (2008) develop in their paper the Munich Chain Ladder (MCL) method to correct the drawbacks of SCL. First, they plot $P_{i,j+1}$ against $Q_{i,j} = P_{i,j}/I_{i,j}$, and find a negative correlation. In plain words, a relative low P/I ratio is followed either by relatively high development factors for paid losses P or relatively low development factors for reported losses and vice versa. It is reasonable since if up to now the paid losses are much less than the reported losses, then much larger paid losses over reported losses ratios must come later, since in the end the paid losses $P_{i,m-1}$ should be equal to the incurred $I_{i,m-1}$.

In the MCL method, regression is used. The development factor $\dfrac{P_{i,j+1}}{P_{i,j}}$ becomes a random variable and its conditional mean (given $Q_{i,j}$) is a first-order polynomial in terms of $Q_{i,j}$, i.e. there exists a constant λ_P such that for all $j = 0, 2, \ldots, m-2$ and all $i = 1, 2, \ldots, m$:

$$\mathbb{E}\Big(\mathbf{Res}\Big(\frac{P_{i,j+1}}{P_{i,j}} \mid \mathcal{P}_i(j)\Big) \mid \mathcal{B}_i(j)\Big) = \lambda_P \cdot \mathbf{Res}(Q_{i,j}^{-1} \mid \mathcal{P}_i(j)), \tag{1.12}$$

where $\mathcal{P}_i(j) := \{P_{i,0}, \ldots, P_{i,j}\}$ stands for the condition that the paid information is given until the end of payment year j for accident year i, $\mathcal{B}_i(j) := \{P_{i,0}, \ldots, P_{i,j}, I_{i,0}, \ldots, I_{i,j}\}$ stands for the knowledge of the development of both processes up to the end of payment year j for accident year i, and the residual is defined as:

$$\mathbf{Res}\Big(\frac{P_{i,j+1}}{P_{i,j}} \mid \mathcal{P}_i(j)\Big) := \frac{\frac{P_{i,j+1}}{P_{i,j}} - \mathbb{E}\Big[\frac{P_{i,j+1}}{P_{i,j}} \mid \mathcal{P}_i(j)\Big]}{\sqrt{\mathbb{V}\Big(\frac{P_{i,j+1}}{P_{i,j}} \mid \mathcal{P}_i(j)\Big)}}. \tag{1.13}$$

Quarg and Mack (2008) also give:

$$\mathbb{E}\Big(\mathbf{Res}\Big(\frac{I_{i,j+1}}{I_{i,j}} \mid \mathcal{I}_i(j)\Big) \mid \mathcal{B}_i(j)\Big) = \lambda_I \cdot \mathbf{Res}(Q_{i,j} \mid \mathcal{I}_i(j)), \text{ for all } i = 1, 2, \ldots, m, \text{ and } j = 0, \ldots, m-2, \tag{1.14}$$

where $\mathcal{I}_i(j) := \{I_{i,0}, \ldots, I_{i,j}\}$ stands for the condition that the incurred development of accident year i is given up to and including j. These mathematical equations are used to model the dependence of the paid and incurred development factors on the preceding (I/P) and (P/I) ratios.

1.2.1.4 Merz-Wüthrich paid incurred chain

Aside from the MCL model, Merz and Wüthrich (2010) present a novel stochastic model for claim reserving that addresses the paid or reported dilemma. They assume that the ratio of any two neighbours in paid chain ladders or reported chain ladders (also called incurred chain ladders) are log normal distributed. Figure 1.2 gives a sketch of the approach. Starting from $P_{i,-1}$ defined as 1, they successively simulate $\xi_{i0}, \xi_{i1}, \ldots, \xi_{im-1}$, alongside calculating $P_{i,0}$, $P_{i,1}, \ldots, P_{i,m-1}$ according to $P_{i,j} = P_{i,j-1}\exp(\xi_{ij})$, finally to reach $P_{i,m-1} = I_{i,m-1}$. The next step is a backward recursion: $I_{i,j-1} = I_{i,j}\exp(-\zeta_{ij-1})$. This model overcomes the problem of SCL where $P_{i,m-1}$ does not equal to $I_{i,m-1}$. Also the dependence between $P_{i,j+1}/P_{i,j}$ and $P_{i,j}/I_{i,j}$ is automatically introduced due to the structure of multivariate normal distribution.

$$\begin{tikzcd}
1 \arrow[r, "\xi_{i0}"] & P_{i,0} \arrow[r, "\xi_{i1}"] & P_{i,1} \arrow[r, "\xi_{i2}"] & \cdots \arrow[r, "\xi_{im-2}"] & P_{i,m-2} \arrow[r, "\xi_{im-1}"] & P_{i,m-1} \arrow[d, equal] \\
 & I_{i,0} & I_{i,1} \arrow[l, "\zeta_{i0}"'] & \cdots \arrow[l, "\zeta_{i1}"'] & I_{i,m-2} \arrow[l, "\zeta_{im-3}"'] & I_{i,m-1} \arrow[l, "\zeta_{im-2}"']
\end{tikzcd}$$

Figure 1.2: PIC Model

1.2.1.5 Double chain ladder

Miranda et al. (2012) derive a simple method for forecasting IBNR and RBNS claims at the same time using the information of reported count data (in a triangular array $\mathcal{N}$) as well as the paid run-off triangle (Δ). Both of these two triangles are usually relatively easy to obtain.

The lifetime of a claim is divided into two: the IBNR delay and the RBNS delay. Unlike most other reserving methods, these two separate sources of delay are estimated separately.

The maximum reporting delay is $m-1$. However, in Miranda et al. (2012), the payment year may exceed $m-1$ due to the settlement delay, see Figure 1.3. Up to the evaluation time, we have the information of aggregated reported counts $\mathcal{N} = \{N_{i,j} \mid i = 1, 2, \ldots, m,\ j = 0, 1, \ldots, m-1\}$ and aggregated payments $\Delta = \{Z_{i,j} \mid i = 1, 2, \ldots, m,\ j = 0, 1, \ldots, m-1\}$.

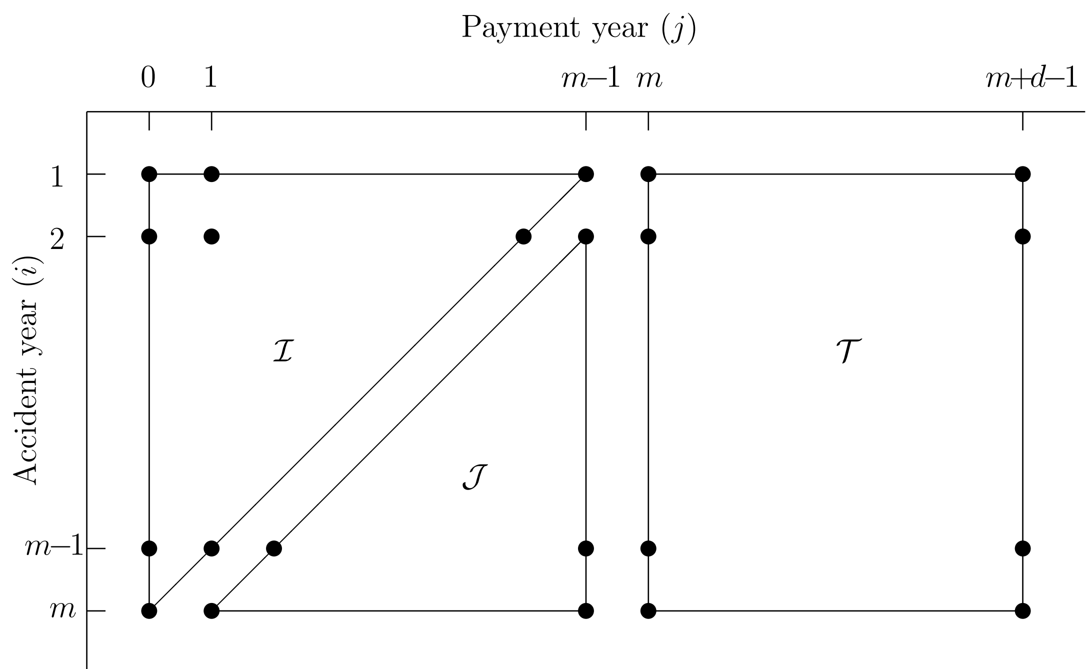

Figure 1.3: Index Sets for Aggregate Claims Data, Maximum Delay Equals d.

The reported counts $N_{i,j}$ are assumed to be independent random variables from a Poisson distribution with multiplicative parametrization:

$$\mathbb{E}[N_{i,j}] = \alpha_i \beta_j, \qquad i = 1, 2, \ldots, m,\ j = 0, 1, \ldots, m-1, \tag{1.15}$$

where $\sum_{j=0}^{m-1} \beta_j = 1$.

A new variable $N_{i,j,l}^{paid}$ is introduced representing the components of $N_{i,j}$ that have settlement delay $SD = l$. Here the conditional distribution of $N_{i,j,l}^{paid}$ given $N_{i,j}$ is supposed to follow a multinomial:

$$(N_{i,j,0}^{paid}, \ldots, N_{i,j,d}^{paid}) \sim \mathbf{Multi}(N_{i,j}; p_0, \ldots, p_d), \qquad i = 1, 2, \ldots, m,\ j = 0, 1, \ldots, m-1, \tag{1.16}$$

where $p = (p_0, p_1, \ldots, p_d)$ denotes the settlement delay probabilities such that $p_0 + p_1 + \cdots + p_d = 1$ with the maximum delay $d \le m - 1$.

In Miranda et al. (2012), it is assumed that for each claim there is only one payment associated with it. The individual payments $Y_{i,j}^{(k)}$ in $Z_{i,j}$ are mutually independent with mean μ_i and variance σ_i^2 such that

$$\mu_i := \mathbb{E}[Y_{i,j}^{(k)}] = \mu\gamma_i \qquad \text{and} \qquad \sigma_i^2 := \mathbb{V}(Y_{i,j}^{(k)}) = \sigma^2\gamma_i^2, \qquad i = 1, 2, \ldots, m, \quad j = 0, 1, \ldots, m-1, \tag{1.17}$$

with μ and σ^2 being mean and variance factors, and i is the inflation over the accident years.

Under the above assumptions the conditional mean of $Z_{i,j}$ becomes

$$\mathbb{E}[Z_{i,j}|\mathcal{N}] = \sum_{l=0}^{\min(j,d)} N_{i,j-l}\, p_l\, \mu\, \gamma_i = \alpha_i\, \mu\, \gamma_i \sum_{l=0}^{\min(j,d)} \beta_{j-l} p_l, \qquad \text{for } i = 1, 2, \ldots, m, \quad \text{and } j = 0, 1, \ldots, m-1. \tag{1.18}$$

Introduce $\tilde{\alpha}_i = \alpha_i\, \mu\, \gamma_i$ and $\tilde{\beta}_j = \sum_{l=0}^{\min(j,d)} \beta_{j-l} p_l$. Then $Z_{i,j}$ has the same multiplicative structure as $N_{i,j}$:

$$\mathbb{E}[Z_{i,j}] = \tilde{\alpha}_i \tilde{\beta}_j, \qquad i = 1, 2, \ldots, m, \; j = 0, 1, \ldots, m-1. \tag{1.19}$$

With these distributional assumptions, the likelihood function can be written as:

$$\mathcal{L}_{\mathcal{N},\Delta} = \mathcal{L}_{\mathcal{N}} \cdot \mathcal{L}_{\Delta|\mathcal{N}}. \tag{1.20}$$

The likelihood function of $\mathcal{N}$ is maximised using the chain ladder method and the other term $\mathcal{L}_{\Delta|\mathcal{N}}$ is approximated using an over-dispersed Poisson distribution. The parameters here are:

1. Delay probabilities: $p_0, \ldots, p_d$.
2. Individual payment parameters: $\mu, \sigma^2, \{\gamma_i \mid i = 1, \ldots, m\}$.
3. Claim counts parameters: α_i, β_j, for $i = 1, \ldots, m$, $j = 0, \ldots, m-1$.

A shortcut is found by applying the standard chain ladder method (CLM) twice to the set $\mathcal{N}$ of $N_{i,j}$ values in (1.15) and to the set Δ of $Z_{i,j}$ values in (1.19) to get the estimates of $\alpha_i, \beta_j, \tilde{\alpha}_i, \tilde{\beta}_j$, then use formula (1.18) to get all the subsequent parameter estimates.

The prediction of RBNS and IBNR reserves is done separately and finally grouped into the forecast of $\hat{Z}_{i,j}$ (for $i + j > m$).

The IBNR component always uses:

$$\hat{Z}_{i,j}^{ibnr} = \sum_{l=0}^{i-m+j-1} \hat{N}_{i,j-l}\,\hat{p}_l\,\hat{\mu}\,\hat{\gamma}_i, \qquad \text{for } i + j > m. \tag{1.21}$$

There are two possible estimates for the RBNS component:

$$\hat{Z}_{i,j}^{rbns(1)} = \sum_{l=i-m+j}^{j} N_{i,j-l}\,\hat{p}_l\,\hat{\mu}\,\hat{\gamma}_i, \qquad \text{for } i + j > m, \tag{1.22}$$

and

$$\hat{Z}_{i,j}^{rbns(2)} = \sum_{l=i-m+j}^{j} \hat{N}_{i,j-l}\,\hat{p}_l\,\hat{\mu}\,\hat{\gamma}_i, \qquad \text{for } i + j > m, \tag{1.23}$$

where $\hat{N}_{i,j} = \hat{\alpha}_i \hat{\beta}_j$. In the cases that $l > d$, then $\hat{p}_l \equiv 0$.

It is shown in Miranda et al. (2012) that with (1.23) for the RBNS component, the estimate of outstanding claims using DCL will be exactly the same as the standard CLM within the non-tail area $\mathcal{J}$:

$$\hat{Z}_{i,j}^{rbns(2)} + \hat{Z}_{i,j}^{ibnr} = \hat{Z}_{i,j}^{CL}. \tag{1.24}$$

However, differences appear when the real count formula in (1.22) is used. Unlike CLM, which only produces forecasts over the region $\mathcal{J}$, DCL also takes into account the tail part $\mathcal{T}$, which is omitted by CLM.

More than the point forecasts of the IBNR and RBNS reserves, Miranda et al. (2011) introduce the bootstrapping procedure to the DCL model for the predictive distributions of the IBNR and RBNS reserves.

A weak point of the DCL method is the lack of stability because the underwriting year inflation near m might be estimated with significant uncertainty. Miranda et al. (2013) propose a model close to DCL but with the inflation γ_i estimated from the less volatile incurred data, then transfering these to the DCL model.

1.2.2 Aggregate Loss Reserving Using GLMs

Wüthrich (2003) applied Tweedie's compound Poisson model, represented as a member of the exponential dispersion family by Jorgensen (1987), to the run-off problem. He defines the model as:

1. The number of payments $R_{i,j}$ in accident year i and payment year j (i.e. cell (i,j)) are independent and Poisson distributed with parameter $\lambda_{i,j}w_i$. The weight $w_i > 0$ is the exposure of each accident year.

2. The individual payments in $R_{i,j}$ are independent and gamma distributed with mean $\tau_{i,j} > 0$ and shape parameter $\gamma > 0$.

3. Denote $Z_{i,j}$ the total incremental payments paid in cell (i,j), and $Y_{i,j} = Z_{i,j}/w_i$.

If we skip the indices i and j. The distribution of Y is parametrized by three parameters λ, τ and γ:

$$f_Y(y;\mu,\phi/w,p) = c(y;\phi/w,p)\exp\left\{\frac{w}{\phi}\left(y\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right\}, \qquad y \geq 0. \tag{1.25}$$

Here the new parameters μ, ϕ and p are chosen to be:

$$\begin{aligned}
&p = (\gamma+2)/(\gamma+1), \qquad p \in (1,2),\\
&\mu = \gamma \cdot \tau,\\
&\phi = \lambda^{1-p}\tau^{2-p}/(2-p).
\end{aligned}$$

A multiplicative model is used to include row and column effects:

$$\mu_{i,j} = \alpha(i) \cdot f(j). \tag{1.26}$$

Compared to the Poisson regression model of Renshaw and Verrall (1998), which is a special case for $p = 1$, it allows more freedom for the distribution of $Z_{i,j}$.

Guszcza and Lommele (2006) also question the traditional methods using only the summarized loss triangles. They point out that these methods can not incorporate the changes in the company's business mix into their estimates of outstanding losses. Another danger of using summarized loss triangles is that they could mask heterogeneous loss development patterns. Finally, traditional methods throw away a large amount of information, prohibiting the use of predictive variables that might determine the loss development. Considering these shortcomings and also for mathematical convenience, Guszcza and Lommele (2006) use a GLM for the development factor in each individual risk class. It is assumed that:

$$\frac{S_{i,j}^{(k)}}{S_{i,j-1}^{(k)}} = \exp(\boldsymbol{\beta}\mathbf{x}^{(k)}) + \sigma, \tag{1.27}$$

where $\mathbf{x}^{(k)}$ represent predictive variables for risk class k and σ is an overdispersed Poisson-distributed error term.

1.2.3 Micro Level Loss Reserving

Lately, a small stream of literature has appeared with a focus on micro-level loss reserving.

In Antonio and Plat (2014) the claim process is treated as a position dependent marked Poisson process. A monthly constant Poisson process is used to model the claim occurrences. A mixture of one Weibull distribution and nine degenerate components are used for the reporting delay. A multiple decrement process defines the development process. The payments are fitted with Burr, gamma, and lognormal distributions with covariate information of initial reserves and the development year for each payment.

Pigeon et al. (2013) suggest the multivariate skew normal distribution for modeling these development factors at a individual claim level.

Let the random variable $Y_{ikj}(>0)$ represent the jth incremental partial amount for the kth claim ($k = 1, \ldots, K_i$) from accident year i ($i = 1, 2, \ldots, I$). Denote by u_{ik} the number of period(s) with partial payment (> 0) after the first one. For a claim (ik) with a strict positive value of u_{ik}, the vector $\Lambda^{ik}_{u_{ik}+1}$ of length $u_{ik} + 1$ gives the development pattern:

$$\boldsymbol{\Lambda}^{(ik)}_{u_{ik}+1} = [Y_{ik1}\ \lambda_1^{(ik)}\ \ldots\ \lambda_{u_{ik}}^{(ik)}]', \tag{1.28}$$

where

$$\lambda_j^{(ik)} = \frac{\sum_{r=1}^{j+1} Y_{ikr}}{\sum_{r=1}^{j} Y_{ikr}}. \tag{1.29}$$

Definition 1.1. *Let $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_k]'$ be a vector of location parameters, $\mathbf{\Sigma}$ a $(k \times k)$ positive definite symmetric scale matrix and $\mathbf{\Delta} = [\Delta_1, \Delta_2, \ldots, \Delta_k]'$ a vector of shape parameters. The $(k \times 1)$ random vector $\mathbf{X}$ follows a multivariate skew-symmetric (MSS) distribution if its density function is of the form*

$$MSS(\mathbf{x}; \mu, \mathbf{\Sigma}^{1/2}, \mathbf{\Delta}) = \frac{2^{(k)}}{det(\mathbf{\Sigma})^{1/2}} g^*(\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \mu)) \prod_{j=1}^{k} H(\Delta_j \mathbf{e}_j' \mathbf{\Sigma}^{-1/2}(\mathbf{x} - \mu)), \tag{1.30}$$

where $g^*(x) = \prod_{j=1}^{k} g(x_j)$, $g(\cdot)$ is a density function symmetric around 0, $H(\cdot)$ is an absolutely continuous cumulative distribution function with $H(\cdot)$ symmetric around 0 and $\mathbf{e}_j'$ are the elementary vectors of the coordinate system $\mathbb{R}^{(k)}$. The MSN distribution is obtained from (1.30) by replacing $g(\cdot)$ and $H(\cdot)$ with the pdf and cdf of the standard normal distribution, respectively.

For closed claims it is assumed that $\ln(\mathbf{\Lambda}_{u_{ik}+1})$ is MSN distributed as:

$$MSN(\ln(\mathbf{\Lambda}_{u_{ik}+1}); \boldsymbol{\mu}_{u_{ik}+1}, \mathbf{\Sigma}_{u_{ik}+1}^{1/2}, \mathbf{\Delta}_{u_{ik}+1} \mid u_{ik}). \tag{1.31}$$

A dependence structure between each partial payment associated with one claim is introduced by this flexible multivariate distribution.

Chapter 2

Full Credibility for GLMMs

In this chapter, we digress temporarily from loss reserve models and study the theory of limited fluctuation credibility for generalized linear mixed models, also called full credibility for GLMMs.

Generalized linear mixed models (GLMMs) are a particular type of mixed model. It is also an extension to the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed effects. It is most widely used for longitudinal data or clustered data analysis. The earliest application of GLMMs in actuarial science could date back to Hachemeister (1975) credibility regression model for U.S. data that showed linear inflation trends in claims. Recently, GLMMs are gaining popularity as a statistical method for insurance data since they combine credibility and GLM for premium rating. The GLMM empirical Bayesian estimator (EBE) is nothing but Bühlmann's Bayesian estimator. In this chapter we extend the theory of limited fluctuation credibility for GLMs to the one for GLMMs. We test the influence of three key factors on the limited fluctuation probability. These are the number of clusters, number of subjects within each cluster and the variance of random effects. Parametric bootstrapping is suggested to derive the limited fluctuation probability with the marginal mean for a general link function.

2.1 Introduction

In actuarial science, Mowbray (1914) first develops a full credibility formula for worker's compensation premium.

If the probability of a small difference between the estimator $\hat{X}$ and the parameter it estimates, m, is "high enough", then the insurer may find $\hat{X}$ credible as an estimator of m. Statistically, this can be defined as

$$\mathbb{P}\{-rm \leq \hat{X} - m \leq rm\} \geq p, \tag{2.1}$$

for a chosen tolerance level $r > 0$ and probability p.

Bühlmann (1967) derives a Bayesian credibility estimators which minimize the square loss function. Jewell (1974) shows that linear credibility estimates are exact when certain natural conjugates are governing the realizations of risk parameters. Hachemeister (1975) worked on U.S data that showed linear inflation trends in claims. This trend differed from one state to the other and also from the average national inflation trend. After a long development of credibility theory, especially in the 60's and 70's, Nelder and Verrall (1997) show how credibility theory can be encompassed within the theory of GLMs. Frees et al. (1999) develop links between credibility theory in actuarial science with longitudinal data models in statistics. They show that many credibility models including Bühlmann, Bühlmann-Straub and the regression model of Hachemeister can be expressed as special cases of the longitudinal data model.

More recently, Zhou and Garrido (2009a) study how the limited fluctuation probability of GLM estimators depend on the sample size, the distribution of covariates and the link function. At the end of their paper, an extension to full credibility for GLMMs is mentioned. However, the formula in Theorem 3.3 of that paper is not strictly a full credibility criterion. We correct this omission and give a real full credibility formula for GLMMs for two quantities: the marginal mean and the cluster specific mean. We show how the limited fluctuation probability depends on the number of clusters, number of subjects within cluster and magnitudes of variance of random effect. Parametric bootstrapping is introduced to simulate the prediction error for the marginal mean.

The outline of this chapter is as follows: Section 2 introduces the notation and numerous computational methods to fit GLMMs. Section 3 shows the application of GLMMs in actuarial science and under which circumstances GLMMs should be preferred over GLMs. Section 4 presents the full credibility results for the marginal mean and also the cluster specific mean of GLMMs. Section 5 shows some numerical experiments we do to inspect the key factors that influence the limited fluctuation probability. Section 6 concludes our work.

2.2 Model and Notation for GLMMs

Generalized linear mixed models (GLMMs) are extensions of both generalized linear models (GLMs) and linear mixed models (LMMs), where the linear predictors contain random effects in addition to the usual fixed effects and the error is not restricted to normal distributed. There are now various books on GLMMs and related topics, see McCullogh and Searle (2001), Demindenko (2004) or Jiang (2007). Antonio and Beirlant (2007) apply GLMMs to estimate and compute several actuarial statistics.

Suppose that data are collected from k different locations or k different years. Each location or year is called a cluster. For the ith cluster we have response data $y_{ij},\ j = 1, 2, ..., n_i$. Let $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ denote the p and q dimension vectors representing fixed effect covariates and random effect covariates associated with the response y_{ij}. Here it is assumed that for each cluster there are random effects $\mathbf{u}_i$ which are added into the regression model to account for the correlation within cluster data. Conditional on $\mathbf{u}_i$, y_{ij} is exponential dispersion distributed with density function of the form:

$$f(y_{ij} \mid \mathbf{u}_i; \boldsymbol{\beta}, \sigma_0^2) = \exp\left\{\frac{w_{ij}}{\sigma_0^2}(y_{ij}\theta_{ij} - b(\theta_{ij})) + c(y_{ij}, \sigma_0^2/w_{ij})\right\}, \qquad y_{ij} \in \mathcal{H}_b, \quad (2.2)$$

where w_{ij} is the weight associated with y_{ij} and the conditional expectation of y_{ij} is:

$$\mathbb{E}(y_{ij}|\mathbf{u}_i; \boldsymbol{\beta}, \sigma_0^2) = b'(\theta_{ij}), \qquad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n_i. \quad (2.3)$$

The regressor is connected with the expectation through a link function g:

$$g(b'(\theta_{ij})) = \eta_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{u}_i, \qquad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n_i. \quad (2.4)$$

The random effects for different clusters are assumed to be i.i.d. random variables distributed as $\pi(\mathbf{u}_i|\mathbf{D})$, usually assumed to be q-variate normal random variables:

$$\mathbf{u}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D}), \qquad \mathbf{D} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2). \tag{2.5}$$

In the following discussion, we will always assume that $\mathbf{D}$ is in diagonal form. As long as the distribution of random effects is completely specified, essentially it makes no difference which distribution is assumed in the development of full credibility.

We use $\boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_q^2)'$ to denote the unknown dispersion parameter σ_0 and those in the diagonal of $\mathbf{D}$, and $\boldsymbol{\psi}$ to denote the unknown parameters for the whole model: $\boldsymbol{\psi} = (\boldsymbol{\beta}', (\boldsymbol{\sigma}^2)')'$.

Generalized linear mixed models can be fitted through maximizing the marginal likelihood function for $\boldsymbol{\psi}$ based on $\boldsymbol{y}$ $(y_{ij},\ i = 1, 2...k,\ j = 1, 2, ...n_i)$:

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} L(\boldsymbol{\psi}|\mathbf{y}) = \arg\max_{\boldsymbol{\psi}} \prod_{i=1}^{k} \Big[\int_{\mathbb{R}^q} [\prod_{j=1}^{n_i} f(y_{ij}|\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0^2)] \pi(\mathbf{u}_i|\mathbf{D}) \mathrm{d}\mathbf{u}_i \Big], \tag{2.6}$$

which involves a multidimensional integration over random effects. Usually these integrals do not have closed-form expressions, with the exception of the normal case.

Various approximation methods have been developed for the ML estimator. The Laplace approximation is one of them, and maybe the earliest one to approximate the likelihood integral. However, Vonesh (1996) shows that under the Laplace approximation:

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(\max\{k^{-\frac{1}{2}}, (\min(n_i))^{-1}\}). \tag{2.7}$$

Intuitively, the $k^{-\frac{1}{2}}$ term comes from standard asymptotic theory while $(\min(n_i))^{-1}$ comes from the Laplace approximation error of the integral. Thus the approximated ML estimator $\hat{\boldsymbol{\beta}}$ will be consistent only when both k and $\min(n_i) \to \infty$.

Almost at the same time, Wolfinger and O'Connell (1993) develop a pseudo-likelihood estimation based on linearization. The advantages of their linearization based method is that it includes a relatively simple linearization form that is well-known and easily fit in linear mixed models (LMMs). Models with correlated errors, a large number of random effects, crossed random effects, and multiple types of subjects can resort to linearization methods. However, the same problem as with the

Laplace approximation arises: the absence of a true objective function for the overall optimization process results in potentially biased estimates, especially for binary data when the number of observations per cluster is small, see Breslow and Lin (1995) and Lin and Breslow (1996).

For these reasons, methods involving Gauss-Hermite quadrature (Liu and Pierce, 1994) and Markov chain Monte Carlo with Gibbs sampling (Zeger and Karim, 1991) have increased in use with the increasing computing power and advancing numerical methods. Both techniques are now available in some SAS and R packages.

After obtaining the estimator $\hat{\boldsymbol{\psi}}$, we can plug it into the joint likelihood function to get empirical posterior mode estimator

$$\hat{\mathbf{u}}_i^{\mathbf{EBM}} = \arg\max_{\mathbf{u}_i} \Big\{ \big(\prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{u}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2) \big) \pi(\mathbf{u}_i | \hat{\mathbf{D}}) \Big\}, \tag{2.8}$$

or empirical posterior mean estimator

$$\hat{\mathbf{u}}_i^{\mathbf{EBE}} = \frac{\displaystyle\int_{\mathbb{R}^q} \mathbf{u}_i \, \big(\prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{u}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2) \big) \, \pi(\mathbf{u}_i | \hat{\mathbf{D}}) \, \mathrm{d}\mathbf{u}_i}{\displaystyle\int_{\mathbb{R}^q} \big(\prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{u}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2) \big) \, \pi(\mathbf{u}_i | \hat{\mathbf{D}}) \, \mathrm{d}\mathbf{u}_i} \,. \tag{2.9}$$

For linear mixed model, the posterior mean estimator equals posterior mode estimator, actually they are well known as best linear unbiased predictor (**BLUP**).

2.3 GLMMs in Actuarial Science

In the most recent decade, GLMMs are gradually adopted in actuarial analysis. GLMMs extend GLMs by including random effects in the linear predictor. The random effects not only determine the correlation structure between subjects in the same cluster, but also take account of heterogeneity among clusters. In this section, we will list some application of GLMMs in actuarial science and the advantages of GLMMs over GLMs one by one.

2.3.1 Bayesian estimator

The actuarial motivation to use GLMMs is that they provide a way of introducing credibility into a generalized linear model setting for ratemaking. It is Bühlmann (1967) that first developed these Bayesian credibility estimators. The Bayesian credibility estimator is the solution that minimize the square loss function.

Klinker (2010) uses GLMs to model the ratio of observed losses to expected losses under current rating plan. He chooses a Tweedie distribution with exponent p between 1 and 2 for this experience ratio. His case study is based on real data from the International Service Office (ISO). In the study, he finds that since there are poorly populated levels in some effects, the standard errors of some estimates are quite larger than for others. Then he applies a GLMM to this data by specify these effects as random effects in SAS PROC GLIMMIX. Note that GLIMMIX should give the estimates in (2.9). After comparing the estimates in these two methods, the evidence of shrinkage to 0 in GLMM estimates compared to GLM estimates is revealed. He also calculates Bühlmann-Straub form credibility estimates at the end of his article. At least in his case study, GLMM estimates are very close to Bühlmann-Straub form credibility estimates.

Jewell (1974) shows that exact credibility occurs (that means that Bühlmann's linear approximation equals the exact Bayesian estimate) when certain natural conjugates are governing the realizations of risk parameters. Thus the Bayesian estimator is a linear combination of exogenous information and sample experience.

Ohlsson and Johansson (2006) extend this result, showing that the exact credibility holds for a class of Tweedie models, including Poisson, gamma and compound Poisson distributions with a special parametrization of random effects.

The Tweedie distribution is generally used in private motor car insurance. The model of the car is an important rating factor, both for third-party liability, hull and theft. Nevertheless, usually we are left with thousands of car models, some of which represent top selling cars with sufficient data available, whereas most classes have moderate or sparse data. In this case, car model can be modelled as random effect in

Tweedie models.

Later on, Ohlsson (2008) demonstrates how to use Bühlmann Straub credibility model in a multiplicative GLM environment and gives the iterative GLMC (GLMs with credibility) algorithm to estimate the parameters.

2.3.2 Efficient estimators

Suppose that we organize data in different clusters, here one cluster could be the data in one territory, in one year or one model of car. If you treat the cluster as a fixed effect, i.e. a categorical variable in GLMs, we may end up drawing good inference about the clusters in the sample. But for a new cluster where we have no experience, we can not set a fair premium for that cluster. In addition, for that new cluster, the value of a random effect would also be a mystery. The marginal mean for that cluster is in the following form:

$$\mu^M = \mathbb{E}[y] = \mathbb{E}\big[\mathbb{E}(y \mid \mathbf{u})\big] = \mathbb{E}\big[g^{-1}(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})\big] = \int g^{-1}(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})\ \pi(\mathbf{u} \mid \mathbf{D})\ \mathrm{d}\mathbf{u}. \tag{2.10}$$

The variance matrix $\mathbf{D}$ of random effects is only acquirable with GLMMs rather than GLMs.

Another solution is simply averaging the estimators of the risk parameters $\mathbf{u}$ in different clusters using GLMs to give the estimator of the marginal mean. This is indeed a solution to the question. The average has the advantage of its unbiasedness and consistency properties. However, it is not an efficient estimator, as is the GLMMs maximum likelihood based estimator.

Recall that for Hachemeister credibility regression model, De Vylder (1981) proves that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{Y} = (\sum_{i=1}^{k}\mathbf{M}_i)^{-1}\sum_{i=1}^{k}\mathbf{M}_i\ \hat{\boldsymbol{\beta}}_i \tag{2.11}$$

has the smallest covariance matrix among all the linear unbiased estimates for $\boldsymbol{\beta}$ (for the definition of matrices $\mathbf{C}$ and $\mathbf{M}$, see Appendix B). Frees et al. (1999) show that Hachemeister credibility regression model can be interpreted as a linear mixed model where $\mathbf{Z}_i = \mathbf{X}_i$. After a simple matrix transformation, we will find that $\hat{\boldsymbol{\beta}}$ in (2.11) is identical to the best linear unbiased estimator (**BLUE**).

In other words, simply averaging the risk parameters in different clusters gives the following estimate:

$$\bar{\boldsymbol{\beta}} = \frac{\sum_{i=1}^{k} \hat{\boldsymbol{\beta}}_i}{k}, \tag{2.12}$$

which is not as efficient as the linear mixed model maximum likelihood estimate $\hat{\boldsymbol{\beta}}$. See Appendix B for the comparison of the covariance matrices of these two estimates using matrix notation.

2.3.3 Correlation and over-dispersion

Another important feature of generalized linear mixed models is that the random effects determine the correlation structure between observations in the same cluster, since they share the same random effects $\mathbf{u}_i$.

We can write the covariance of y_{ij_1} and y_{ij_2} as two parts:

$$\begin{aligned} \mathrm{COV}(y_{ij_1}, y_{ij_2}) &= \mathrm{COV}[\mathbb{E}(y_{ij_1}|\mathbf{u_i}), \mathbb{E}(y_{ij_2}|\mathbf{u_i})] + \mathbb{E}[\mathrm{COV}(y_{ij_1}, y_{ij_2}|\mathbf{u_i})] \\ &= \mathrm{COV}[\mathbb{E}(y_{ij_1}|\mathbf{u_i}), \mathbb{E}(y_{ij_2}|\mathbf{u_i})], \end{aligned} \tag{2.13}$$

therefore GLMMs allow for dependence among these observations. Note that in the GLMs, $\mathrm{COV}(y_{ij_1}, y_{ij_2}) = 0$, the observations in the same cluster are independent.

For one observation y_{ij_1},

$$\mathbb{V}(y_{ij_1}) = \mathrm{COV}[\mathbb{E}(y_{ij_1}|\mathbf{u_i})] + \mathbb{E}[\mathbb{V}(y_{ij_1}|\mathbf{u_i})] \geq \mathbb{E}[\mathbb{V}(y_{ij_1}|\mathbf{u_i})]. \tag{2.14}$$

We can see the inclusion of a random effect introduces over-dispersion. From the above two points in this section, we can see that GLMMs have wider applications in practice than GLMs.

2.3.4 Incidental parameter problem

Lastly, GLMMs are good solutions for the incidental parameter problem of GLMs that usually brings asymptotic inconsistency.

The incidental parameter problem is typically seen to arise with longitudinal data models when the cluster specific intercepts are allowed in a regression model. The problem is that maximum-likelihood estimates of the structural parameters need not be consistent.

There are two famous examples in the literature:

Example 1. Let y_{ij} be distributed as:

$$f(y_{ij}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(y_{ij} - u_i)^2/2\sigma^2\}, \qquad i = 1, 2, \dots, k, \; j = 1, 2, \dots, n.$$

u_i is the cluster specific intercept. If we treat it as a linear regression model with different intercept for each cluster, then maximum likelihood estimates of u_i and σ^2 are:

$$\hat{u}_i = \bar{y}_i,$$
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)}{kn}.$$

It is known that $\hat{\sigma}^2 \sim \frac{\sigma^2 \chi^2_{k(n-1)}}{kn}$ with expectation $\sigma^2(n-1)/n$. When $k \to \infty$, but n is constant, we see that the estimators of σ^2 are not consistent. The bias is not mitigated by increasing the number of clusters.

From a Bayesian point of view, it suggests a way of thinking about the construction of a reasonable prior for the intercept u_i. It is easy to see that if the intercepts u_i are i.i.d. and normally distributed, then the maximum likelihood estimators of σ^2 in this linear mixed model is consistent with the number of clusters.

Example 2. In a binary data model, suppose y to be binary with:

$$\mathbb{E}(y_{ij}) = \text{logit}(u_i + \beta x_{ij}), \qquad i = 1, 2, \dots, k, \; j = 1, 2, \dots, n.$$

For these models the maximum likelihood estimates is generally inconsistent as $k \to \infty$ for β. For example, when $n = 2, \; x_{i1} = 0, \; x_{i2} = 1$ then $\hat{\beta} \to 2\beta$ (Andersen (1970)). However, if u_i are i.i.d. and normally distributed, then the maximum likelihood estimates of β in this logit-normal model are consistent.

See Lancaster (2000) for a detailed survey of the history of incidental parameter problems in statistics and in the econometrics literature.

2.4 Full Credibility for GLMMs

Zhou and Garrido (2009a) study the limited fluctuation probability of GLM estimators. At the end of their paper, an extension to full credibility for GLMMs is mentioned.

They calculate the following variance function:

$$\mathbb{V}(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}-\eta_{ij}) = \mathbb{V}(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}-\mathbf{x}'_{ij}\boldsymbol{\beta}-\mathbf{z}'_{ij}\mathbf{u}_i) = \mathbb{V}(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}})+\mathbb{V}(\mathbf{z}'_{ij}\mathbf{u}_i) = \mathbf{x}'_{ij}\boldsymbol{\Omega}\mathbf{x}_{ij}+\mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}, \quad (2.15)$$

$\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}$ is the predictor for the fixed effect part. The linear component $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{u}_i$ is a random variable since it includes $\mathbf{u}_i$. $\boldsymbol{\Omega} = \mathrm{COV}(\hat{\boldsymbol{\beta}})$ is the variance-covariance matrix of the fixed-effects parameter estimator $\hat{\boldsymbol{\beta}}$. $\mathbf{D}$ is the covariance parameter for random effects, see equation (2.5).

However, in terms of credibility, we are not looking at the mean square difference between a predictor and a random variable. What we really care about is the probability that an estimate falls into a small region around the fixed (the marginal mean) or the realized value (the cluster specific mean). Then we can see how efficient and credible the estimate is, and what the crucial factors are.

Note that $\mathbb{V}(\mathbf{z}'_{ij}\mathbf{u}_i) = \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}$, a value that does not converge to 0 with increasing sample size, so using the formula in Zhou and Garrido (2009a) we cannot reach the accepted truth that the limited fluctuation probability converges to 0 as observations increase.

Next we give a more proper credibility criterion for the marginal mean and also for the cluster specific mean.

2.4.1 Full credibility for marginal means

For insurance practice, we need to set fair premium for the next year based on previous years data. An essential and indispensable value is the expectation of future losses. If future losses depend on the realization of the random effect, then the total expectation (marginal mean) for the subject $(\mathbf{x}, \mathbf{z})$ is expressed as the following integral:

$$\mu^M = \mathbb{E}[y] = \mathbb{E}\big[\mathbb{E}(y \mid \mathbf{u})\big] = \mathbb{E}\big[g^{-1}(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})\big] = \int_{\mathbb{R}^q} g^{-1}(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})\ \pi(\mathbf{u} \mid \mathbf{D})\ \mathrm{d}\mathbf{u}, \tag{2.16}$$

here $\mathbf{x}$, $\mathbf{z}$ are column vectors representing the covariates for the subject we considered.

2.4.1.1 The log link function

Lemma 2.1. *The log-link function which is widely used in claim frequency and severity modeling to prevent the appearance of negative estimators, whereby μ^M has a closed form formula:*

$$\begin{aligned} \mu^M &= \int_{\mathbb{R}^q} \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})\ (2\pi)^{-\frac{q}{2}}|\mathbf{D}|^{-\frac{1}{2}}\ \exp(-\frac{1}{2}\mathbf{u}'\mathbf{D}^{-1}\mathbf{u})\mathrm{d}\mathbf{u} \\ &= \exp(\mathbf{x}'\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}'\mathbf{D}\mathbf{z}). \end{aligned} \tag{2.17}$$

The integral here is similar to the moment generating function for a multivariate normal distribution or the expectation of a log-normal distribution.

Lemma 2.2. *Use $\hat{\mu}^M = \exp(\mathbf{x}'\hat{\boldsymbol{\beta}} + \frac{1}{2}\mathbf{z}'\hat{\mathbf{D}}\mathbf{z})$ as the estimator of μ^M. Then for a chosen tolerance level $r > 0$, the limited fluctuation probability*

$$\begin{aligned} \pi &= \mathbb{P}\{|\hat{\mu}^M - \mu^M| \le r\mu^M\} = \mathbb{P}\{(1-r)\mu^M \le \hat{\mu}^M \le (1+r)\mu^M\} \\ &= \mathbb{P}\{\ln[(1-r)\mu^M] \le \ln(\hat{\mu}^M) \le \ln[(1+r)\mu^M]\} \\ &= \mathbb{P}\{\ln(1-r) \le (\mathbf{x}'\hat{\boldsymbol{\beta}} + \frac{1}{2}\mathbf{z}'\hat{\mathbf{D}}\mathbf{z}) - (\mathbf{x}'\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}'\mathbf{D}\mathbf{z}) \le \ln(1+r)\} \\ &= \mathbb{P}\{\ln(1-r) \le \mathbf{x}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{1}{2}\mathbf{z}'(\hat{\mathbf{D}} - \mathbf{D})\mathbf{z} \le \ln(1+r)\} \end{aligned} \tag{2.18}$$

Lemma 2.3. *The ML estimator $\hat{\boldsymbol{\psi}}$ is a consistent and asymptotic normally distributed estimator of $\boldsymbol{\psi}$:*

$$\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \sim \mathbf{N}(\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\psi})), \tag{2.19}$$

where $\mathcal{I}(\boldsymbol{\psi})$ is the fisher information matrix associated with the likelihood function in (2.6).

Booth and Hobert (1998) use this in an logistic-normal example. Jiang (2007) is a good reference for the asymptotic properties of different inference approaches to GLMMs.

Theorem 2.1. *If we denote by*

$$\mathbf{m} = \begin{pmatrix} \mathbf{x} \\ 0 \\ \frac{1}{2}\,\mathbf{z} \circ \mathbf{z} \end{pmatrix},$$

where $\mathbf{z} \circ \mathbf{z}$ is the element-wise product of $\mathbf{z}$ with itself, also known as Hadamard product, and 0 is the coefficient for σ_0 which is not used in the estimate μ^M, then we have that:

$$\mathbf{x}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{1}{2}\mathbf{z}'(\hat{\mathbf{D}} - \mathbf{D})\mathbf{z} \quad \approx \quad N(0, \mathbf{m}'\,\mathcal{I}^{-1}(\boldsymbol{\psi})\,\mathbf{m}). \tag{2.20}$$

Denote by $s = (\mathbf{m}'\,\mathcal{I}^{-1}(\boldsymbol{\psi})\,\mathbf{m})^{\frac{1}{2}}$, put it back into (2.18), and finally we get:

$$\pi \approx \Phi(\frac{\ln(1+r)}{s}) - \Phi(\frac{\ln(1-r)}{s}), \tag{2.21}$$

where Φ is the cumulative distribution function of standard normal variable, and ϕ is the corresponding probability density function.

2.4.1.2 The probit link function

Lemma 2.4. *For the probit link function ($g^{-1} = \Phi$) and normally distributed random effects ($\mathbf{D}$ is not necessarily diagonal), we can simplify the integral in (2.16) as follows:*

$$\int_{\mathbb{R}^q} \Phi(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})(2\pi)^{-\frac{q}{2}}|\mathbf{D}|^{-\frac{1}{2}}\,\exp(-\frac{1}{2}\mathbf{u}'\mathbf{D}^{-1}\mathbf{u})\,\mathrm{d}\mathbf{u} = \Phi\Big(\frac{\mathbf{x}'\boldsymbol{\beta}}{\sqrt{1+\mathbf{z}'\mathbf{D}\mathbf{z}}}\Big), \tag{2.22}$$

which is the multi-dimensional version of the following result in the normal integral, see Owen (1980):

$$\int_{-\infty}^{\infty} \Phi(a+bx)\;\phi(x)\;\mathrm{d}x = \Phi\Big(\frac{a}{\sqrt{1+b^2}}\Big). \tag{2.23}$$

Lemma 2.5. *Use* $\hat{\mu}^M = \Phi\Big(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}}{\sqrt{1+\mathbf{z}'\hat{\mathbf{D}}\mathbf{z}}}\Big)$ *as the estimator of* μ^M *for the probit link function. Then for chosen tolerance level* $r > 0$*, the limited fluctuation probability*

$$\begin{aligned}\pi &= \mathbb{P}\{|\hat{\mu}^M - \mu^M| \leq r\mu^M\} \\ &= \mathbb{P}\{(1-r)\mu^M \leq \hat{\mu}^M \leq (1+r)\mu^M\} \\ &= \mathbb{P}\Big\{(1-r)\Phi\Big(\frac{\mathbf{x}'\boldsymbol{\beta}}{\sqrt{1+\mathbf{z}'\mathbf{D}\mathbf{z}}}\Big) \leq \Phi\Big(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}}{\sqrt{1+\mathbf{z}'\hat{\mathbf{D}}\mathbf{z}}}\Big) \leq (1+r)\Phi\Big(\frac{\mathbf{x}'\boldsymbol{\beta}}{\sqrt{1+\mathbf{z}'\mathbf{D}\mathbf{z}}}\Big)\Big\}.\end{aligned} \tag{2.24}$$

It is hard to get a closed-form expression for π with the probit link function. However, parametric bootstrapping can help us solve this problem. The implementation of parametric bootstrapping is quite similar to the nonparametric bootstrapping, the only difference is that instead of simulating bootstrapped samples that are i.i.d. from the empirical distribution, we choose to simulate bootstrap samples that are i.i.d. from the estimated parametric model. This technique is especially useful in a regression framework, since there we do not have i.i.d. samples. Since the estimates of regression coefficients and other scale parameters are available, finally we can give any bootstrapped quantity of interest.

We start with the parametric bootstrapping method for limited fluctuation probability in GLMs, the notation here is the same to that in Zhou and Garrido (2009a).

1. Based on the sample data $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$, get the coefficient estimator $\hat{\boldsymbol{\beta}}$ and dispersion parameter $\hat{\phi}$. Plug $\hat{\boldsymbol{\beta}}$ into the mean estimator $\hat{\mu}_i = g^{-1}(\mathbf{X}_i'\hat{\boldsymbol{\beta}})$.

2. Simulate samples $\mathbf{Y}^k = \{Y_1^k, \ldots, Y_n^k\}$ from the exponential dispersion distributions with the parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$.

3. Recalculate the estimator $\hat{\boldsymbol{\beta}}^k$ from the new sampling $\mathbf{Y}^k$, plug it into $\hat{\mu}_i^k = g^{-1}(\mathbf{X}_i'\hat{\boldsymbol{\beta}}^k)$ and check the inequality $|\hat{\mu}_i^k - \hat{\mu}_i| \leq r\hat{\mu}_i$. If it is hold, then $\zeta_k = 1$, otherwise $\zeta_k = 0$.

4. Repeat step 2 to 3 for $k = 1, 2, \ldots, m$.

5. Estimate π_i by $\hat{\pi}_i = \dfrac{1}{m}\displaystyle\sum_{k=1}^{m}\zeta_k$.

This algorithm can be greatly simplified if we apply the asymptotic distribution of β^k which is the normal distribution $\mathbf{N}(\boldsymbol{\beta}, \mathcal{I}^{-1}(\boldsymbol{\beta}))$. Finally we will get the same result as Theorem 3.2 in Zhou and Garrido (2009a).

Next we apply bootstrapping to formula (2.24):

1. Based on the given sample data $\mathbf{Y} = \{y_{ij} \mid i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i\}$, obtain the maximum likelihood estimator $\hat{\boldsymbol{\psi}}$ and the observed variance-covariance function $\mathcal{I}^{-1}(\hat{\boldsymbol{\psi}})$.

2. Generate $\hat{\boldsymbol{\psi}}^k$ from the normal distribution $\mathbf{N}(\hat{\boldsymbol{\psi}}, \mathcal{I}^{-1}(\hat{\boldsymbol{\psi}}))$.

3. Check the inequality $(1-r)\Phi\Big(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}^k}{\sqrt{1+\mathbf{z}'\hat{\mathbf{D}}^k\mathbf{z}}}\Big) \le \Phi\Big(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}}{\sqrt{1+\mathbf{z}'\hat{\mathbf{D}}\mathbf{z}}}\Big) \le (1+r)\Phi\Big(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}^k}{\sqrt{1+\mathbf{z}'\hat{\mathbf{D}}^k\mathbf{z}}}\Big)$, if the two-sided inequality holds, then $\zeta_k = 1$, otherwise $\zeta_k = 0$.

4. Repeat step 1 to 2 m times.

5. Estimate π by $\hat{\pi} = \dfrac{1}{m}\displaystyle\sum_{k=1}^{m} \zeta_k$.

2.4.1.3 The general link function

The logit link function is widely used in the modeling of binary data. If we want to give a limited fluctuation probability in this case, then the difficult task is the evaluation of the following integral to determine the inequality $|\hat{\mu}^M - \mu^M| \le r\mu^M$ is true or false. Note that this expression is the mean for the logit normal distribution, and there is no analytical solution for it:

$$\int\limits_{\mathbb{R}^q} \frac{\exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})} (2\pi)^{-\frac{q}{2}} |\mathbf{D}|^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{u}'\mathbf{D}^{-1}\mathbf{u}) d\mathbf{u} \tag{2.25}$$

The good news is that we can simplify this multivariate integral to a univariate integral, since $\mathbf{z}'\mathbf{u}$ enters into the logit function as a unit, we simplify it as:

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}) &= \int\limits_{\mathbb{R}^q} g^{-1}(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{u})(2\pi)^{-\frac{q}{2}} |\mathbf{D}|^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{u}'\mathbf{D}^{-1}\mathbf{u}) d\mathbf{u} \\ &= \int\limits_{-\infty}^{\infty} g^{-1}(\mathbf{x}'\boldsymbol{\beta} + y)\ (2\pi)^{-\frac{1}{2}} |\mathbf{z}'\mathbf{D}\mathbf{z}|^{-\frac{1}{2}} \exp\Big(-\frac{y^2}{2\mathbf{z}'\mathbf{D}\mathbf{z}}\Big) dy\ , \end{aligned} \tag{2.26}$$

thus significantly reducing the computation burden.

Finally, we summarize the parametric bootstrapping method to develop a full credibility criterion for general link function:

1. Based on the given sample data $\mathbf{Y} = \{y_{ij} \mid i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i\}$, obtain the maximum likelihood estimator $\hat{\boldsymbol{\psi}}$ and the observed variance-covariance function $\mathcal{I}^{-1}(\hat{\boldsymbol{\psi}})$.

2. Generate $\hat{\boldsymbol{\psi}}^k$ from the normal distribution $\mathbf{N}(\hat{\boldsymbol{\psi}}, \mathcal{I}^{-1}(\hat{\boldsymbol{\psi}}))$.

3. Evaluate $L(\hat{\boldsymbol{\beta}}^k, \hat{\mathbf{D}}^k)$ and check the inequality $|L(\hat{\boldsymbol{\beta}}^k, \hat{\mathbf{D}}^k) - L(\hat{\boldsymbol{\beta}}, \hat{\mathbf{D}})| \leq r\, L(\hat{\boldsymbol{\beta}}, \hat{\mathbf{D}})$. If it holds, then $\zeta_k = 1$, otherwise $\zeta_k = 0$.

4. Repeat Steps 1 to 2 m times.

5. Estimate π by $\hat{\pi} = \dfrac{1}{m}\displaystyle\sum_{k=1}^{m} \zeta_k$.

For log link GLMMs, Theorem 2.1 will save the work of generating $\hat{\boldsymbol{\psi}}^k$ as well as the evaluation of the integral (2.16) for each generated random $\hat{\boldsymbol{\psi}}^k$. In terms of probit link GLMMs, it is hard to avoid simulations to generate $\hat{\boldsymbol{\psi}}^k$, but Lemma 2.4 saves the trouble of evaluating the integral in (2.16).

Proposition 2.1. *The limited fluctuation probability π in (2.21) is a generalization of the one in Theorem 3.1 of Zhou and Garrido (2009a). If the model specification is that there are no random effects, i.e. $\mathbf{D} = 0$, then π in (2.21) is exactly the same to that in (3.11) of Zhou and Garrido (2009a). Our full credibility criterion is applicable for GLM model, random coefficient model (Bayesian GLM), random intercept model and random effect model (generic GLMM).*

2.4.2 Full credibility for cluster specific means

We know the random variables in GLMMs consist of two parts $\mathbf{Y} = \{y_{ij} \mid i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i\}$ and $\mathbf{U} = \{\mathbf{u}_i \mid i = 1, 2, \ldots, k\}$. Here $\mathbf{Y}$ are observed while $\mathbf{U}$ are latent variables.

The cluster i specific mean for subject $(\mathbf{x}_{ij}, \mathbf{z}_{ij})$ is defined as:

$$\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i). \tag{2.27}$$

We introduce two notation $\hat{\mathbf{u}}_i$ and $\hat{\hat{\mathbf{u}}}_i$, where:

$$\hat{\mathbf{u}}_i = \arg\max_{\mathbf{u}_i} L(\mathbf{u}_i \mid \mathbf{y}_i, \boldsymbol{\psi}) = \arg\max_{\mathbf{u}_i} \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{u}_i, \boldsymbol{\beta}, \sigma_0^2). \tag{2.28}$$

and

$$\hat{\hat{\mathbf{u}}}_i = \arg\max_{\mathbf{u}_i} L(\mathbf{u}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\psi}}) = \arg\max_{\mathbf{u}_i} \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{u}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2), \tag{2.29}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are estimated through (2.6).

Fitting $\mathbf{u}_i$ with given $\boldsymbol{\psi}$ could be done using the GENMOD procedure in SAS with $\mathbf{y}_i$ as the observation, $\mathbf{u}_i$ as the regression coefficient, $\mathbf{z}_{ij}$ as the covariate and $\mathbf{x}'_{ij}\boldsymbol{\beta}$ as the offset-that is, a regression variable with a constant coefficient of 1 for each observation.

After all we can give an estimator for the cluster specific mean by the following:

$$\hat{\eta}_{ij} = \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} + \mathbf{z}'_{ij}\hat{\hat{\mathbf{u}}}_i. \tag{2.30}$$

We do not use the empirical Bayesian estimator to derive the full credibility for cluster specific mean estimators since we can not have a compact formula with full credibility for EBE. Also for a fixed $\mathbf{u}_i$, asymptotically distributions of EBE and EBM converge to the distribution for $\hat{\hat{\mathbf{u}}}_i$ with large n_i.

Note that our estimator $\hat{\eta}_{ij}$ in (2.30) is a function on $\mathbf{Y}$ only. However, the target value η_{ij} which we want to see how credible the estimator $\hat{\eta}_{ij}$ is has latent variables $\mathbf{U}$ included.

We use $\mathcal{P}$ to denote the probability space of $(\mathbf{Y}, \mathbf{U})$. Since $\hat{\eta}_{ij}$ is a function of $\mathbf{Y}$, it is also a random variable which we could denote as $\hat{\eta}_{ij}(\mathbf{Y})$. Generally speaking, η_{ij} is still a random variable, since it includes $\mathbf{U}$, and more precisely $\mathbf{u}_i$, thus we can denote it by $\eta_{ij}(\mathbf{u}_i)$.

In Figure 2.1 we use the set on left side to denote the probability space $\mathcal{P}$. $(\mathbf{Y}, \mathbf{U})$ is an element in this space. It is mapped to two points on the real line $\mathbb{R}$, $\hat{\eta}_{ij}(\mathbf{Y})$

and $\eta_{ij}(\mathbf{u}_i)$ respectively. $\hat{\eta}_{ij}(\mathbf{Y})$ is known while $\eta_{ij}(\mathbf{u}_i)$ is unknown, so we use dashed arrow pointing to $\eta_{ij}(\mathbf{u}_i)$.

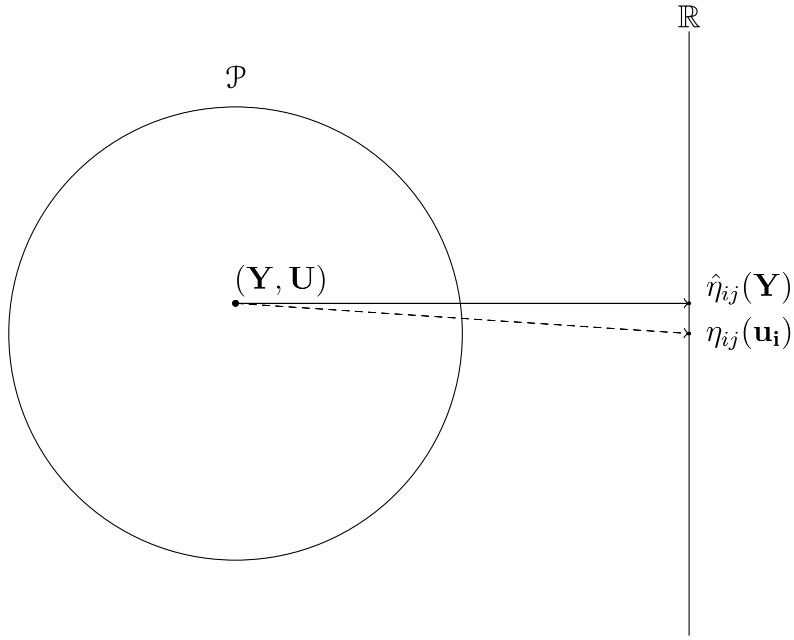

Figure 2.1: GLMM Random Variable and Estimate

A problem of credibility arises for our estimator $\hat{\eta}_{ij}$. In other words, in which circumstance we could adopt $\hat{\eta}_{ij}$ as an estimator of η_{ij} with an acceptable error.

We suggest the conditional limited fluctuation probability $\mathbb{P}\{|\hat{\mu}_{ij} - \mu_{ij}| \leq r\mu_{ij} \mid \mathbf{u}_i\}$ as a credibility criterion. In Figure 2.2, a subset in the probability space $\mathcal{P}$ is defined as $\mathbf{S} = \{(\mathbf{Y}, \mathbf{U}) \mid \mathbf{u}_i = \mathbf{u}_i^0\}$, where $\mathbf{u}_i^0$ is the value of the random effect for cluster i in the given sample. It is easy to see that each element of this subset should generally map to different $\hat{\eta}_{ij}$, but they all map to the same η_{ij}, that is $\eta_{ij}(\mathbf{u}_i^0)$.

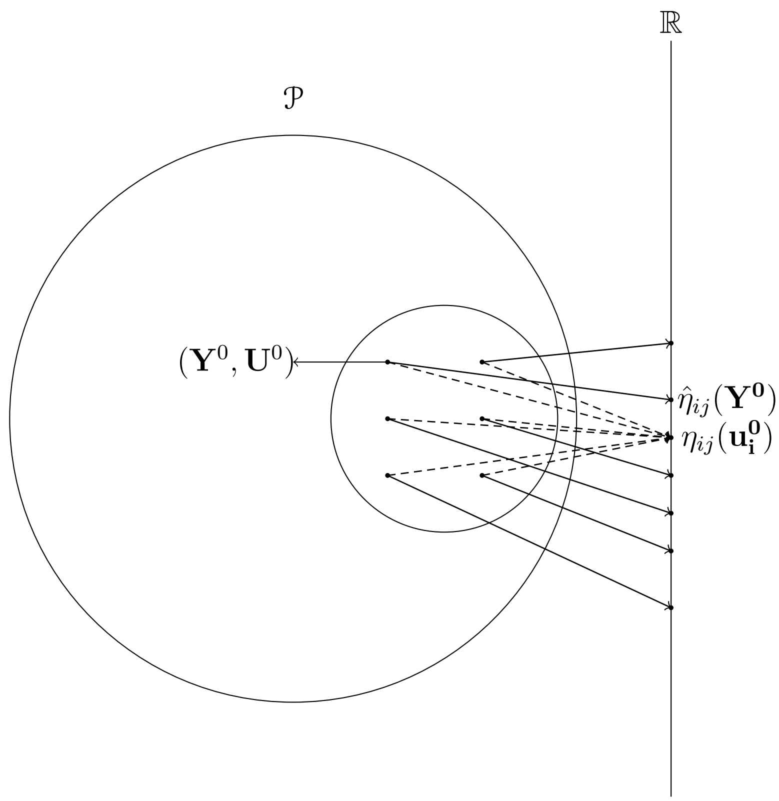

Figure 2.2: GLMM Random Variable Subset and Estimates

From the figure above we see that $\mathbb{P}\{|\hat{\mu}_{ij} - \mu_{ij}| \leq r\mu_{ij} \mid \mathbf{u}_i\}$ is a measure of dispersion for the image set $\hat{\eta}_{ij}(\mathbf{S})$ in the real line. In other words, we have no idea about the exact distance between $\hat{\eta}_{ij}$ and η_{ij} in the given data set, but we do know that our sample is an element for the subset $\mathbf{S}$ and luckily it is feasible to give the dispersion rate of $\hat{\eta}_{ij}(\mathbf{S})$ in $\mathbb{R}$.

Lemma 2.6. *Let g be a monotonic increasing link function. Then $\mu_{ij} = \mathbb{E}[y_{ij} \mid \mathbf{u}_i] = g^{-1}(\eta_{ij}) = g^{-1}(\mathbf{x}_{ij}^t\boldsymbol{\beta} + \mathbf{z}_{ij}^t\mathbf{u}_i)$. The conditional probability*

$$\begin{aligned}
\pi_{ij}(\mathbf{u}_i) &= \mathbb{P}\{|\hat{\mu}_{ij} - \mu_{ij}| \leq r\mu_{ij} \mid \mathbf{u}_i\} = \mathbb{P}\{(1-r)\mu_{ij} \leq \hat{\mu}_{ij} \leq (1+r)\mu_{ij} \mid \mathbf{u}_i\} \\
&= \mathbb{P}\{g[(1-r)\mu_{ij}] - g(\mu_{ij}) \leq g(\hat{\mu}_{ij}) - g(\mu_{ij}) \leq g[(1+r)\mu_{ij}] - g(\mu_{ij}) \mid \mathbf{u}_i\} \\
&= \mathbb{P}\{g[(1-r)\mu_{ij}] - \eta_{ij} \leq \hat{\eta}_{ij} - \eta_{ij} \leq g[(1+r)\mu_{ij}] - \eta_{ij} \mid \mathbf{u}_i\}. \qquad (2.31)
\end{aligned}$$

In practice, $g[(1-r)\mu_{ij}] - \eta_{ij}$ and $g[(1+r)\mu_{ij}] - \eta_{ij}$ are replaced by their estimated values. For a log link function, we can simplify these to $\log(1-r)$ and $\log(1+r)$ as found in Zhou and Garrido (2009a).

Theorem 2.2. *Asymptotically, $\hat{\eta}_{ij} - \eta_{ij}$ could be conditionally approximated, given $\mathbf{u}_i$, by:*

$$\hat{\eta}_{ij} - \eta_{ij} \approx N(0, \mathbf{A}'_{ij}\mathbf{\Omega}\mathbf{A}_{ij} + \mathbf{z}'_{ij}(\mathbf{Z}'_i\mathbf{W}_0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij}). \tag{2.32}$$

The technical derivations can be found in Appendix A. Similarly to Theorem 2.1, let

$$s_{ij} = (\mathbf{A}'_{ij}\mathbf{\Omega}\mathbf{A}_{ij} + \mathbf{z}'_{ij}(\mathbf{Z}'_i\mathbf{W}_0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij})^{\frac{1}{2}},$$

and for a log link function, insert this into (2.31) to get

$$\pi_{ij}(\mathbf{u}_i) \approx \Phi\Big(\frac{\ln(1+r)}{s_{ij}}\Big) - \Phi\Big(\frac{\ln(1-r)}{s_{ij}}\Big). \tag{2.33}$$

Proposition 2.2. *When the fixed effect covariates $\mathbf{X}_i$ is identical to the random effect coefficient design matrix $\mathbf{Z}_i$, then in (2.32) we have $\mathbf{A}_{ij} = \mathbf{0}$, and $\mathbf{A}'_{ij}\mathbf{\Omega}\mathbf{A}_{ij} = 0$, therefore $s_{ij} = (\mathbf{z}'_{ij}(\mathbf{Z}'_i\mathbf{W}_0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij})^{\frac{1}{2}} = (\mathbf{x}'_{ij}(\mathbf{X}'_i\mathbf{W}_0\mathbf{X}_i)^{-1}\mathbf{x}_{ij})^{\frac{1}{2}}$. It is easy to see that our limited fluctuation probability in (2.33) is exactly the same to the one if we take the responses in the ith cluster out, and apply the GLM credibility criteria in Zhou and Garrido (2009a) to those data. Note that when $\mathbf{X}_i = \mathbf{Z}_i$, this is a random coefficient model. The implication of this proposition is that for a random coefficient model, all the information useful to estimate the cluster specific mean bears in the cluster itself.*

2.4.3 Unconditional mean square error of prediction

The conditional limited fluctuation probability $\pi_{ij}(\mathbf{u}_i)$ is a criterion to determine whether we should use $\hat{\eta}_{ij}$ for η_{ij} in this given sample. Not like full credibility with GLM, since it is conditional on the random effect $\mathbf{u}_i$, it is inappropriate to use this criterion again with the same covariates but for a different sample. Unconditional Mean Square Error of Prediction (UMSEP) seems to be good solution for this problem.

Definition 2.1. UMSEP *is defined as the expectation of the square error between our predictor (estimator) $\hat{\eta}_{ij}$ and η_{ij} (the expectation of y_{ij}):*

$$\text{UMSEP} \;\equiv\; \mathbb{E}\big[\big(\hat{\eta}_{ij} - \eta_{ij}\big)^2\big]. \tag{2.34}$$

Lemma 2.7. *Using the tower property of conditional expectation, we can write* UMSEP *as:*

$$\text{UMSEP} \;\equiv\; \mathbb{E}\big[\big(\hat{\eta}_{ij} - \eta_{ij}\big)^2\big] \;=\; \mathbb{E}\Big[\mathbb{E}\big[\big(\hat{\eta}_{ij} - \eta_{ij}\big)^2 \;\big|\; \mathbf{u}_i\big]\Big]. \tag{2.35}$$

If we replace $\mathbb{E}\big[\big(\hat{\eta}_{ij}-\eta_{ij}\big)^2 \mid \mathbf{u}_i\big]$ with the asymptotic conditional variance $\mathbf{A}'_{ij}\boldsymbol{\Omega}\mathbf{A}_{ij} + \mathbf{z}'_{ij}$ $(\mathbf{Z}'_i\mathbf{W}^0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij}$ in Theorem 2.32, then we have:

$$\int\limits_{\mathbb{R}^q} \big(\mathbf{A}'_{ij}\boldsymbol{\Omega}\mathbf{A}_{ij} + \mathbf{z}'_{ij}\,(\mathbf{Z}'_i\mathbf{W}^0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij}\big)\;(2\pi)^{-\frac{q}{2}}|\mathbf{D}|^{-\frac{1}{2}}\;\exp(-\frac{1}{2}\mathbf{u}'_i\mathbf{D}^{-1}\mathbf{u}_i)\;\mathrm{d}\mathbf{u}_i. \tag{2.36}$$

In Section 2.5.3 we apply a Gauss-Hermite Quadrature to this integral and compare the approximation with Monte Carlo simulations.

2.5 Numerical Illustration

In this section we use a simulated Poisson-normal GLMM to detect the key factors that influence the full credibility for the marginal mean and the cluster specific means.

In this example, the $\mathbf{R}$ side covariates (fixed effect covariates) are set to be $\mathbf{x}'_{ij} = (1 \quad b_{ij})$, where the covariate b_{ij} is simulated from the normal distribution $N(1, 0.25^2)$. Fixed effect coefficient $\boldsymbol{\beta}' = (\beta_0 \quad \beta_1)$ are chosen to be $(1 \quad 1)$. The $\mathbf{G}$ side effect includes only the random intercept for each cluster, which is assumed to be distributed as $N(0, \sigma_1^2)$.

After all the covariates and parameters are fixed, we first simulate a random intercept effect:

$$u_i \quad \sim \quad N(0, \sigma_1^2), \qquad \text{for } i = 1, 2, \ldots, k. \tag{2.37}$$

Then we continue to simulate the response y_{ij} as:

$$y_{ij} \quad \sim \quad Pois(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ij} + u_i)), \qquad \text{for } i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n_i. \tag{2.38}$$

2.5.1 Marginal mean

The GLIMMIX procedure in SAS/STAT (2006) is a useful package. The Hessian matrix is included in the output when we specify the “H” option. The covariance matrix

$$\mathcal{I}^{-1}(\boldsymbol{\psi}) = \begin{pmatrix} \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_0) & \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{cov}(\hat{\beta}_0, \hat{\sigma}_1^2) \\ \mathrm{cov}(\hat{\beta}_1, \hat{\beta}_0) & \mathrm{cov}(\hat{\beta}_1, \hat{\beta}_1) & \mathrm{cov}(\hat{\beta}_1, \hat{\sigma}_1^2) \\ \mathrm{cov}(\hat{\sigma}_1^2, \hat{\beta}_0) & \mathrm{cov}(\hat{\sigma}_1^2, \hat{\beta}_1) & \mathrm{cov}(\hat{\sigma}_1^2, \hat{\sigma}_1^2) \end{pmatrix} \tag{2.39}$$

is calculated by the observed inverse Fisher information matrix, which equals $2\mathbf{H}^{-1}$.

First we fix the number of clusters $k = 10$, $\sigma_1 = 1$ and allow the number of subjects within clusters to vary. Using (2.20), we give $2\mathbf{H}^{-1}$ and the limited fluctuation probability for the subject with fixed coefficient $(1 \quad 1)$:

$$n_i = 20, \qquad \begin{pmatrix} 0.0590021 & -0.01049 & -0.000753 \\ -0.01049 & 0.0098112 & -0.000016 \\ -0.000753 & -0.000016 & 0.0461208 \end{pmatrix}, \qquad \pi = 0.3050573.$$

$$n_i = 200, \qquad \begin{pmatrix} 0.0524326 & -0.000452 & -0.000057 \\ -0.000452 & 0.0004264 & -1.71E-6 \\ -0.000057 & -1.71E-6 & 0.0540886 \end{pmatrix}, \qquad \pi = 0.2930831.$$

$$n_i = 2000, \qquad \begin{pmatrix} 0.0551134 & -0.000099 & -0.000017 \\ -0.000099 & 0.0000936 & 1.8257E-7 \\ -0.000017 & 1.8257E-7 & 0.0605255 \end{pmatrix}, \qquad \pi = 0.282379.$$

We find that prediction error for x_0 and σ_1 can not be reduced when we increase the number of subjects within clusters and full credibility is not assigned to the marginal mean.

Next we keep the n_i at 20, set $k = 1000$ and give the new $2\mathbf{H}^{-1}$:

$$\begin{pmatrix} 0.0010455 & -0.000073 & -0.000019 \\ -0.000073 & 0.0000685 & -1.957E-7 \\ -0.000019 & -1.957E-7 & 0.0019003 \end{pmatrix}, \qquad \pi = 0.9915864.$$

Prediction error is negligible and limited fluctuation probability is already at 99%. Intuitively, as more clusters are included, more experience is added about the distribution of the random effect, estimation of the random effect variance parameter σ_1

should be more accurate. Finally it will contribute to the precision of the marginal mean estimator through formula (2.17).

Then we keep $k = 10, n_1 = 20$, and compare the results with different σ_1:

$$\sigma_1 = 1, \qquad \begin{pmatrix} 0.0590021 & -0.01049 & -0.000753 \\ -0.01049 & 0.0098112 & -0.000016 \\ -0.000753 & -0.000016 & 0.0461208 \end{pmatrix}, \qquad \pi = 0.3050573.$$

$$\sigma_1 = 0.1, \qquad \begin{pmatrix} 0.0140964 & -0.011708 & 0.0000476 \\ -0.011708 & 0.0110104 & -0.000064 \\ 0.0000476 & -0.000064 & 0.0000523 \end{pmatrix}, \qquad \pi = 0.9846788.$$

We find that the less variation the random effect has, the more credible our marginal mean estimator is.

2.5.2 The cluster specific mean

Now we turn to the conditional limited fluctuation probability for the cluster specific mean of the subject $\mathbf{x}'_{11} = (1 \quad 1)$. We fix the random effect $u_1 = 0$ and $n_i = 10$ ($i = 1, \ldots, k$), vary the number of clusters,

	$k = 2,\ n_i = 10$	$k = 20,\ n_i = 10$	$k = 200,\ n_i = 10$	$k = 2000,\ n_i = 10$	$k = 2,\ n_i = 200$
$\hat{\pi}_{11}(u_1)$	0.6377517	0.6596258	0.6714299	0.6882975	0.9995769

Table 2.1: Cluster Specific Mean and n_i

We find that increasing the number of clusters could increase $\hat{\pi}_{11}(u_1)$ to a certain extent but this stops at some upper-limit. The reason is that increasing the number of clusters will eliminate the first variance component in (2.32) but has no influence on the second component.

However, increasing the number of subjects within cluster lets $\hat{\pi}_{11}(u_1)$ increase to 1. Since we will have enough experience in each cluster, intuitively we should trust the cluster specific mean estimator $\hat{\eta}_{11}$. Statistically we find that both variance components in (2.32) diminish when we increase n_1 to infinity.

We could also explain this phenomena in plain words. Since the linear predictor $\hat{\eta}_{11}$ includes $\boldsymbol{\beta}$ and u_1, adding more clusters will offer more experience only for $\boldsymbol{\beta}$. But adding more subjects in cluster 1 will provide more experience to infer about both $\boldsymbol{\beta}$ and u_1.

Next we show the extent to which the conditional limited fluctuation probability $\pi_{11}(u_1)$ is determined by u_1. We set $k = 50, n_i = 20$ and besides $\hat{\pi}_{11}(u_1)$ we also give $\pi^*_{11}(u_1)$ from the Monte Carlo simulations.

u_1	-3.0000	-2.0000	-1.0000	0.0000	1.0000	2.0000	3.0000
η_{11}	-1.0000	0.0000	1.0000	2.0000	3.0000	4.0000	5.0000
$\hat{\eta}_{11}$	-1.1700	-0.1689	0.9001	1.9342	3.1041	3.9806	5.0146
$\hat{\mathbb{V}}(\hat{\eta}_{11} \mid u_1)$	0.1848	0.0583	0.0223	0.0059	0.0018	0.0008	0.0003
$\hat{\pi}_{11}(u_1)$	0.1845	0.3222	0.4981	0.8076	0.9812	0.9995	0.9999
$\pi^*_{11}(u_1)$	0.2720	0.3281	0.5210	0.8062	0.9661	0.9997	0.9999

Table 2.2: Conditional Variance and Random Intercept

Figure 2.3 plots η_{11} and $\hat{\eta}_{11}$ against u_1. From the picture we see that when the random intercept u_1 takes negative values near -3, the difference between our estimator $\hat{\eta}_{11}$ and the real value of η_{11} is relatively large. And as the random intercept increases, $\hat{\eta}_{11}$ and η_{11} converge.

Therefore, we can say that the random effect on which our conditional limited fluctuation probability in (2.31) is conditioned has a large impact on $\hat{\pi}_{11}(u_1)$.

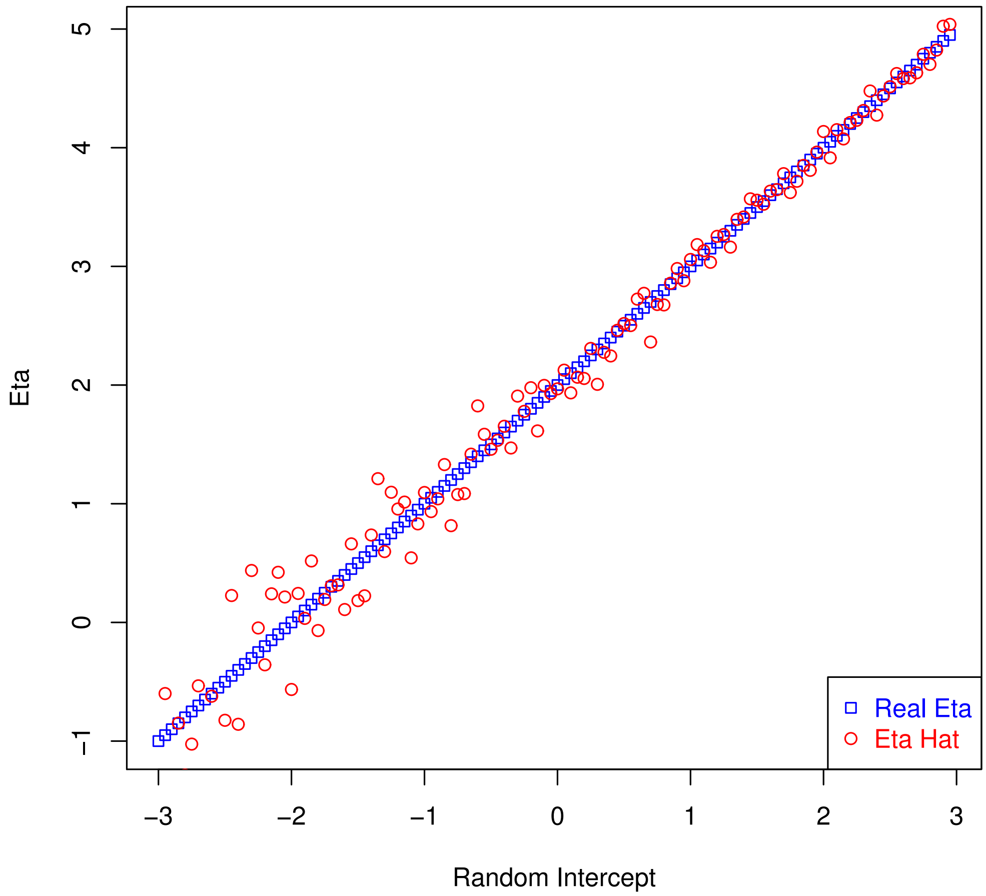

Figure 2.3: η_{11} versus $\hat{\eta}_{11}$

2.5.3 Numerical approximation for UMSEP

Let UMSEP* denote the UMSEP calculated by integrating the conditional variance over u_1, and UMSEP* represent the estimates of UMSEP using Monte Carlo simulations with 10,000 iterations. Table 2.1 shows the results for different k and n_i.

	$k=20,\ n_i=20$	$k=50,\ n_i=20$	$k=50,\ n_i=40$
UMSEP*	0.011531	0.010366	0.005731
UMSEP*	0.012247	0.010958	0.005979

Table 2.3: Integration over Conditional Variance vs Monte Carlo Simulation

We find that our algorithm for deriving UMSEP gives a good performance and saves significant computation time compared to Monte Carlo simulations.

2.6 Summary

This chapter discusses the motivations to use GLMMs beyond GLMs. The inclusion of random effects in the linear predictor not only offers the correlation structure within clusters but also provides one way to combine credibility rate making with GLMs.

The important definitions and concepts on model formulation, numerical estimation and prediction are summarized.

Then we study the credibility of the marginal mean and cluster specific mean estimators obtained from generalized linear mixed risk models. Some closed forms of full credibility criteria are given as well as a parametric bootstrapping algorithm for approximating the limited fluctuation probability.

In addition to the sample size, the number and distribution of the covariates and the link function that are noticed previously in the study of full credibility for GLMs, the random effect is a new decisive factor in the full credibility for GLMMs.

Most of our work is based on previous research, we generalize their results to the random effects case.

Chapter 3

Individual Loss Reserving Using GLMs and GLMMs

As we said in Chapter 1, most of the loss reserving methods are based on aggregate loss development triangle and were created in an age when the computing power was expensive. Each method has its advantages and disadvantages but cannot all be applied simultaneously. These methods are not fully adequate to capture the complexities of the stochastic reserving for general insurance. Verdonck et al. (2009) illustrate that the outstanding claims reserves by the CLM are strongly affected by outliers. Another potential danger is that they mask the heterogeneous loss development patterns for different risk classes. Fuchs (2014) shows that generally the CLM estimates applied to full portfolios are different from the sum of CLM estimates applied to sub portfolios. It seems that only a method based on individual risk class level data could incorporate the changes in the company's mix of business into the estimates of outstanding losses.

Zhou and Garrido (2009b) establish a complete structural reserving method on individual risk class level. They incorporate loss emergence and development patterns, connecting frequency and severity, and embed them all in the framework of GLMs. In this chapter, we refine their original work. First, several parametric functions are used to fit the claim reporting delay in an interval censored and right truncated regression model, similar to the method in Rosenberg (1990). Then we proceed to the loss emergence mechanism, using a Poisson regression model. Since real life count data are frequently characterized by over-dispersion and excess zeros, we also apply the zero-inflated negative binomial regression model. After that we estimate the future reported claim numbers for each risk class. Claim severity and claim settlement delays are modeled next. Finally, we give the estimates of the IBNR losses and RBNS losses separately for each risk class, similarly to the method used in Miranda et al. (2012), then add them up to get the estimates for the total loss reserve in the overall business.

3.1 Reporting Delay

3.1.1 Truncated data and interval censored data

Let Y denote a random response variable, and let y denote its observed value, $T^{(l)}$ and $T^{(r)}$ denote the random variables for the left-truncation and right-truncation threshold respectively, and let $t^{(l)}$ and $t^{(r)}$ denote their realized values for one observation.

If there is no left-truncation, then $t^{(l)} = \tau_l$, where τ_l is the smallest value in the support of some given response distribution, so $F(t^{(l)}) = 0$; similarly if there is no right-truncation, then $t^{(r)} = \tau_h$, where τ_h is the largest value in the support of the distribution, so $F(t^{(r)}) = 1$.

Let $C^{(l)}$ and $C^{(r)}$ denote the random variables for the left-censoring and right-censoring limit, respectively, and let $c^{(l)}$ and $c^{(r)}$ denote their values for an observation, respectively. If there is no left-censoring, then $c^{(l)} = \tau_h$, so $F(c^{(l)}) = 1$. If there is no right-censoring, then $c^{(r)} = \tau_l$, so $F(c^{(r)}) = 0$.

In SAS/STAT (2009), the set of input observations can be categorized into the following four subsets:

1. E is the set of uncensored and untruncated observations. The likelihood of an observation in E for a response that has a parametric distribution F_Θ with corresponding density f_Θ is

$$l = Pr(Y = y) = f_\Theta(y).$$

2. E_t is the set of uncensored observations that are truncated. The likelihood of an observation in E_t is

$$l_{E_t} = Pr(Y = y | t^{(l)} < Y \leq t^{(r)}) = \frac{f_\Theta(y)}{F_\Theta(t^{(r)}) - F_\Theta(t^{(l)})}.$$

3. C is the set of censored observations that are not truncated. The likelihood of an observation in C is

$$l_c = Pr(c^{(r)} < Y \leq c^{(l)}) = F_\Theta(c^{(l)}) - F_\Theta(c^{(r)}).$$

4. C_t is the set of censored observations that are truncated. The likelihood of an observation in C_t is

$$l_{C_t} = Pr(c^{(r)} < Y \leq c^{(l)} | t^{(l)} < Y \leq t^{(r)}) = \frac{F_\Theta(c^{(l)}) - F_\Theta(c^{(r)})}{F_\Theta(t^{(r)}) - F_\Theta(t^{(l)})}.$$

3.1.2 Distribution with scale parameter

We fit the reporting delay (RD) by a distribution family $\mathcal{F}(\Theta)$ which has a scale parameter or log-transformed scale parameter Θ. If the regression effects are not modelled, then the distribution for response random variable RD is assumed to be:

$$RD \sim \mathcal{F}(\Theta_0),$$

for a particular "true" scale parameter Θ_0. We add regression effects to the distribution, then the shape of distribution of RD is spread out or compressed according to:

$$RD \sim \exp\big(\sum \boldsymbol{\beta}_r' \mathbf{x}_r^{(k)}\big) \mathcal{F}(\Theta_0).$$

If Θ is a scale parameter for the family $\mathcal{F}$, then

$$\exp\big(\sum \boldsymbol{\beta}_r' \mathbf{x}_r^{(k)}\big) \mathcal{F}(\Theta_0) = \mathcal{F}\Big(\exp\big(\sum \boldsymbol{\beta}_r' \mathbf{x}_r^{(k)}\big) \Theta_0\Big).$$

Table 3.1 gives the predefined distribution that could be used by the SEVERITY procedure in SAS/ETS (2010b).

Name	Distribution	Parameters	Pdf (f) and Cdf (F)
BURR	Burr	$\theta > 0, \alpha > 0$	$f(x) = \frac{\alpha\gamma z^{\gamma}}{x(1+z^{\gamma})^{(}\alpha+1)}$
		$\gamma > 0$	$F(x) = 1 - \left(\frac{1}{1+z^{\gamma}}\right)^{\alpha}$
EXP	Exponential	$\theta > 0$	$f(x) = \frac{1}{\theta}e^{-z}$
			$F(x) = 1 - e^{-z}$
GAMMA	Gamma	$\theta > 0, \alpha > 0$	$f(x) = \frac{z^{\alpha}e^{-z}}{x\Gamma(\alpha)}$
			$F(x) = \frac{\gamma(\alpha,z)}{\Gamma(\alpha)}$
GPD	Generalized	$\theta > 0, \xi > 0$	$f(x) = \frac{1}{\theta}(1+\xi z)^{-1-1/\xi}$
	Pareto		$F(x) = 1 - (1+\xi z)^{-1/\xi}$
IGAUSS	Inverse Gaussian	$\theta > 0, \alpha > 0$	$f(x) = \frac{1}{\theta}\sqrt{\frac{\alpha}{2\pi z^3}}e^{\frac{-\alpha(z-1)^2}{2z}}$
	(Wald)		$F(x) = \Phi\left((z-1)\sqrt{\frac{\alpha}{z}}\right) + \Phi\left(-(z+1)\sqrt{\frac{\alpha}{z}}\right)e^{2\alpha}$
LOGN	Lognormal	μ (no bounds),	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$
		$\sigma > 0$	$F(x) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$
PARETO	Pareto	$\theta > 0, \alpha > 0$	$f(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}$
			$F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$
WEIBULL	Weibull	$\theta > 0, \tau > 0$	$f(x) = \frac{1}{x}\tau z^{\tau}e^{-z^{\tau}}$
			$F(x) = 1 - e^{-z^{\tau}}$

Notes:

1. $z = x/\theta$, wherever z is used.
2. θ denotes the scale parameter for all the distributions. For LOGN, $\log(\theta) = \mu$.
3. Parameters are listed in the order in which they are defined in the distribution model.
4. $\gamma(a,b) = \int_0^b t^{a-1}e^{-t}dt$ is the lower incomplete gamma function.
5. $\Phi(y) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{y}{\sqrt{2}}\right)\right)$ is the standard normal cdf.

Table 3.1: Parametric Function for Reporting Delay

3.1.3 Likelihood function for reporting delay parameters

Denote the likelihood function for reporting delay parameters $\boldsymbol{\beta}_r$ as $L_r(\boldsymbol{\beta}_r \mid \mathcal{N})$, we can write it as the product of separate likelihoods for each risk class:

$$L_r(\boldsymbol{\beta}_r \mid \mathcal{N}) = \prod_{k=1}^{K} L_r(\boldsymbol{\beta}_r \mid \mathcal{N}^{(k)}). \tag{3.1}$$

The contribution to $L_r(\boldsymbol{\beta}_r \mid \mathcal{N}^{(k)})$ due to one reported claim in $N_{i,j}^{(k)}$ is:

$$\frac{F(j+1\ ;\ \theta_k) - F(j\ ;\ \theta_k)}{F(m+1-i\ ;\ \theta_k)}, \tag{3.2}$$

where $\theta_k = \exp\left(\sum \boldsymbol{\beta}_r' \mathbf{x}_r^{(k)}\right) \cdot \theta_0$, and θ_0 is usually inside an exponential function as the intercept.

Therefore combining with Table 3.2, we know that:

$$L_r(\boldsymbol{\beta}_r \mid \mathcal{N}^{(k)}) = \prod_{i=1}^{m} \prod_{j=0}^{m-i} \left(\frac{F(j+1\ ;\ \theta_k) - F(j\ ;\ \theta_k)}{F(m+1-i\ ;\ \theta_k)} \right)^{N_{i,j}^{(k)}}. \tag{3.3}$$

Accident	Reporting Delay								
Year	0	1	$\cdots$	j	$\cdots$	$m-i$	$\cdots$	$m-2$	$m-1$
1	$N_{1,0}^{(k)}$	$N_{1,1}^{(k)}$	$\cdots$	$N_{1,j}^{(k)}$	$\cdots$	$N_{1,m-i}^{(k)}$	$\cdots$	$N_{1,m-2}^{(k)}$	$N_{1,m-1}^{(k)}$
2	$N_{2,0}^{(k)}$	$N_{2,1}^{(k)}$	$\cdots$	$N_{2,j}^{(k)}$	$\cdots$	$N_{2,m-i}^{(k)}$	$\cdots$	$N_{2,m-2}^{(k)}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$			
i	$N_{i,0}^{(k)}$	$N_{i,1}^{(k)}$	$\cdots$	$N_{i,j}^{(k)}$	$\cdots$	$N_{i+1,m-i}^{(k)}$			
$\vdots$	$\vdots$	$\vdots$		$\vdots$					
$m-j$	$N_{m-j,0}^{(k)}$	$N_{m-j,1}^{(k)}$	$\cdots$	$N_{m-j,j}^{(k)}$					
$\vdots$	$\vdots$	$\vdots$							
$m-1$	$N_{m-1,0}^{(k)}$	$N_{m-1,1}^{(k)}$							
m	$N_{m,0}^{(k)}$								

Table 3.2: Claim Count Run-off Triangle

Other than the ready-made SEVERITY procedure in SAS, also the optimization subroutines of the IML procedure in SAS/IML (2010) can be used to maximize the likelihood function $L_r(\boldsymbol{\beta}_r \mid \mathcal{N})$ for parameters $\boldsymbol{\beta}_r$. These two methods give the same result, but the interactive matrix language (IML) give us more freedom to specify even some piecewise functions for reporting delay.

Based on the parameter estimates $\hat{\boldsymbol{\beta}}_r$, we know that $\hat{\theta}^{(k)} = \exp\left(\sum \hat{\boldsymbol{\beta}}'_r \mathbf{x}_r^{(k)}\right) \cdot \theta_0$ and the estimates of $p_j^{(k)}$ can be written as:

$$\hat{p}_j^{(k)} = F(j+1\ ;\ \hat{\theta}^{(k)}) - F(j\ ;\ \hat{\theta}^{(k)}), \tag{3.4}$$

and

$$\hat{p}_{(i)j}^{(k)} = \frac{F(j+1\ ;\ \hat{\theta}^{(k)}) - F(j\ ;\ \hat{\theta}^{(k)})}{F(m+1-i\ ;\ \hat{\theta}^{(k)})}, \tag{3.5}$$

where $p_j^{(k)}$ and $p_{(i)j}^{(k)}$ are the individual level p_j and $p_{(i)j}$, see Rosenberg (1990).

3.2 Claim Numbers

3.2.1 Poisson regression model

In Zhou and Garrido (2009b), a Poisson regression model is used to fit the claim counts for different accident years. The overall incurred claim number in accident year i and risk class k is assumed to be Poisson distributed with mean $w_i^{(k)} \exp(\boldsymbol{\beta}'_f \mathbf{x}_f^{(k)})$. A log link function is used here to prevent negative numbers.

Lemma 3.1. *In addition to the Poisson assumption for claim count, if we assume that the reporting delay for each claim is independent, then we know $N_{i,j}^{(k)}$ are independently distributed as:*

$$N_{i,j}^{(k)} \sim Pois(w_i^{(k)} \exp(\boldsymbol{\beta}'_f \mathbf{x}_f^{(k)}) p_j^{(k)}), \qquad \text{for } i = 1, 2, \ldots, m \text{ and } j = 0, 1, \ldots, m-1. \tag{3.6}$$

In Zhou and Garrido (2009b), the parametric function F chosen for reporting delay is either EXP or WEIBULL (see Table 3.1), $\hat{p}_j^{(k)}$ is then calculated according to (3.4). Based on Lemma 3.1, the likelihood function for $\boldsymbol{\beta}_f$ is:

$$L(\boldsymbol{\beta}_f\ ;\ \mathcal{N},\ \hat{p}) = \prod_{k=1}^{K} L(\boldsymbol{\beta}_f\ ;\ \mathcal{N}^{(k)},\ \hat{p}^{(k)}) = \prod_{k=1}^{K} \prod_{i=1}^{m} \prod_{j=0}^{m-i} L(\boldsymbol{\beta}_f\ ;\ N_{i,j}^{(k)},\ \hat{p}_j^{(k)}), \tag{3.7}$$

here $\hat{p} = \{\hat{p}_j^{(k)} \mid k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, m-1\}$ and $\hat{p}^{(k)} = \{\hat{p}_j^{(k)} \mid j = 0, 1, \ldots, m-1\}$. $L(\boldsymbol{\beta}_f\ ;\ N_{i,j}^{(k)},\ \hat{p}_j^{(k)})$ is calculated based on (3.6).

There are also two methods to get the estimates of $\boldsymbol{\beta}_f$, either by nonlinear optimization or using the GENMOD procedure in SAS/STAT (2013). Note here if we resort to the GENMOD procedure, $\ln(w_i^{(k)} p_j^{(k)})$ is set as the offset.

Estimates of INBR claim numbers are given by:

$$\hat{N}_{i,j}^{(k)} = w_i^{(k)} \exp(\hat{\boldsymbol{\beta}}_f' \mathbf{x}_f^{(k)}) \hat{p}_j^{(k)}, \qquad i + j > m. \tag{3.8}$$

This method seems to be the simplest method when applying GLMs to individual level loss reserves and could be implemented easily with any statistical software.

3.2.2 Negative binomial and zero-inflated model

In some occasions, claim count data may not follow the usual Poisson distribution, in particular if they are zero-inflated and over dispersed. The number of observed zeros may exceed the number of expected zeros under the Poisson or the negative binomial distribution assumptions.

We apply a zero-inflated negative binomial model to the claim frequency data since it includes the Poisson model, negative binomial model, zero-inflated Poisson model.

Let y denote the number of claims incurred in one accident year for one policyholder. The assumption of zero inflated negative binomial regression model is:

$$\mathbf{ZINB}(y\ ;\ \pi_0, \lambda, \phi) = \begin{cases} \pi_0 + (1 - \pi_0)\mathbf{NB}(y\ ;\ \lambda, \phi), & \text{if } y = 0, \\ (1 - \pi_0)\mathbf{NB}(y\ ;\ \lambda, \phi)\ , & \text{if } y \in \mathbb{Z}^+. \end{cases} \tag{3.9}$$

$\pi_0 = \pi_0(\boldsymbol{\gamma}'\mathbf{z})$ is called zero-inflated link function. It relates $\boldsymbol{\gamma}'\mathbf{z}$, the multiplication of zero inflated covariates vector $\mathbf{z}$ and regression coefficient $\boldsymbol{\gamma}$ to the probability of excess zeros. Usually it is set as either a logistic function or the standard normal cumulative distribution function (the probit function).

NB is negative binomial distribution as:

$$\mathbf{NB}(y\ ;\ \lambda, \phi) = \frac{\Gamma(y + \phi)}{y!\Gamma(\phi)} \Big(\frac{\phi}{\phi + \lambda}\Big)^{\phi} \Big(\frac{\lambda}{\phi + \lambda}\Big)^{y}, \qquad y \in \mathbb{N}, \tag{3.10}$$

where usually the mean λ is linked to $\boldsymbol{\beta}'_f \mathbf{x}_f$ through:

$$\lambda = \exp(\boldsymbol{\beta}'_f \mathbf{x}_f).$$

Lemma 3.2. *The mean and variance of the zero inflated negative binomial model are:*

$$\begin{aligned} \mathbb{E}[y] &= \lambda(1-\pi_0), & (3.11) \\ \mathbb{V}[y] &= \lambda(1-\pi_0)(1+\lambda(\pi_0+\phi)). & (3.12) \end{aligned}$$

Notice that zero inflated negative binomial model exhibits overdispersion if at least one of two parameters π_0 and ϕ are greater than 0*.*

For the fitting of zero inflated models, we use the "COUNTREG" procedure in SAS, see SAS/ETS (2010a).

See Appendix C for the use of the zero inflated negative binomial model for claim frequency and how to give the estimates for the future reported claim numbers.

3.2.3 Hurdle model

The hurdle model is another interesting alternative to Poisson and negative binomial models for the analysis of claims reported by an insured. It includes the zero inflated model as well as models with less zeros than expected.

There are two processes controlling the hurdle model. The basic idea is that firstly a Bernoulli probability governs the binary outcome of whether the count variate is a zero or positive realization. Then if the realization is positive, the hurdle is crossed, and the conditional distribution of the positive values is governed by a truncated-at-zero count data model. Thus the hurdle model is formulated as:

$$\mathbf{H}(y) = \begin{cases} \pi_0, & \text{if } y = 0, \\ (1-\pi_0)\dfrac{f(y)}{1-f(0)}, & \text{if } y \in \mathbb{Z}^+. \end{cases} \tag{3.13}$$

Lemma 3.3. *The hurdle negative binomial distribution is defined as:*

$$\mathbf{HNB}(y\ ;\ \pi_0, \lambda, \phi) = \begin{cases} \pi_0, & \textit{if } y = 0, \\ (1-\pi_0)\dfrac{\mathbf{NB}(y\ ;\ \lambda, \phi)}{1-\mathbf{NB}(0\ ;\ \lambda, \phi)}, & \textit{if } y \in \mathbb{Z}^+, \end{cases} \tag{3.14}$$

where $\mathbf{NB}(y\ ;\ \lambda, \phi)$ *denotes the negative binomial distribution as in (3.10).*

It could also be defined as a mixture of hurdle Poisson distributions:

$$\mathbf{HPois}(y\ ;\ \pi_0, \lambda, \tau) = \begin{cases} \pi_0, & \textit{if } y = 0, \\ (1-\pi_0)\dfrac{\mathbf{Pois}(y\ ;\ \lambda\tau)}{1-\mathbf{Pois}(0\ ;\ \lambda\tau)}, & \textit{if } y \in \mathbb{Z}^+, \end{cases} \tag{3.15}$$

where $\mathbf{Pois}(y\ ;\ \lambda\tau)$ *denotes the Poisson distribution with rate equal to* $\lambda\tau$ *and the prior distribution for* τ *is:*

$$g(\tau\ ;\ \phi, \lambda) = \frac{1-\exp(-\lambda\tau)}{1-\left(\frac{\phi}{\phi+\lambda}\right)^{\phi}} \frac{\phi^{\phi}}{\Gamma(\phi)}\tau^{\phi-1}\exp(-\phi\tau), \qquad \tau > 0. \tag{3.16}$$

See Appendix D for the use of the hurdle negative binomial model for claim frequency and how to give the estimates for the future reported claim numbers.

Boucher et al. (2007) present and compare different risk classification models for the annual number of claims reported to the insurer. They choose the best distribution describing the data based on several specification tests for nested or non-nested models and goodness-of-fit test.

3.2.4 Accident year effect

In the Poisson model for claim numbers, the Poisson rate for the same risk class is assumed to be constant over the accident years. However, it is possible that there are some accident year effects which could inflate or deflate the expected claim numbers.

Note that in the double chain ladder model, a chain ladder method is applied to the aggregated reported claim counts triangle $\mathcal{N}$. Since the chain ladder method is equivalent to a Poisson regression model with accident years and reporting delays as categorical covariates, thus, there are m levels $\alpha_1, \alpha_2, ...\alpha_m$ for accident year effects, where α_m is often set to 0 to prevent multicollinearity.

However, since $\mathcal{N}$ is a triangle, there are few data in the last few rows, therefore there will be larger standard errors for the estimates of the accident year effect in last few accident years than the estimates for the first few accident years. This is a problem as the projection of future reserve is composed mostly of the entries in the last few rows.

In other words, the introduce of the accident year effect may reduce the unbiasedness, but it will increase the variance of the reserve estimations. Instead of blindly applying a chain ladder method to the reported claim triangle $\mathcal{N}$, a statistical hypothesis test on $\alpha_i = 0$ $(i = 1, 2, \ldots, m-1)$ should be done first. See Appendix F for the simulation study on this issue.

Another way to grasp the year effect is to assume that for each accident year there is a common prior for the random year effect u_i.

Gigante et al. (2013) assume a hierarchical overdispersed Poisson model for the incremental payments with gamma distributed risk parameters.

The overdispersed Poisson-gamma model estimated through the h-likelihood approach provides, for each origin year, a reserve estimate that is a mixture of two reserves: one based on the run-off data, the other based on external data.

The advantage of the random effect model is that when the data is too scarce to be credible in the last few rows, the random effect model will give a Bayesian credibility estimate.

3.3 Future Reported Numbers

For future reported numbers, i.e. the numbers of the incurred but not reported claims, we only need the reporting delay model and the claim frequency model. The first step is to choose the best parametric distribution in Table 3.1, and get the estimators $\hat{p}_j^{(k)}$. Then we treat the estimates $\hat{p}_j^{(k)}$ as the real values, plug them into the claim frequency modeling process. Finally we give the estimates of $\hat{N}_{i,j}^{(k)}$ $(i + j > m)$ for each risk class.

3.4 Claim Severity

For claim severity (total payments associated with one claim), Zhou and Garrido (2009b) use a gamma GLM. Probability distributions of the response Y are parameterized as follows, see SAS/STAT (2009):

$$f(y) = \frac{1}{\Gamma(v)y}\Big(\frac{yv}{\mu}\Big)^{v} \exp\Big(-\frac{yv}{\mu}\Big), \tag{3.17}$$

where $\mathbb{E}[Y] = \mu$, $\mathbb{V}[Y] = \dfrac{\mu^2}{v}$ and $\phi = v^{-1}$. Through a link function g, the expected response $\mu = \mathbb{E}[Y]$ is related to the linear predictor $\eta = \boldsymbol{\beta}'\mathbf{x}$, that is $g(\mu) = \eta$. Here v is the "scale" parameter displayed in the standard output of the GENMOD procedure in SAS.

In the fitted claim severity process, each payment is discounted first by an inflation index before they are added up for each closed (settled) claim, then this sum Y is used as the response. For each risk class, $\mu^{(k)}$ is assumed to be related to the explanatory variables $\mathbf{x}_s^{(k)}$ through:

$$\mu^{(k)} = \exp(\boldsymbol{\beta}_s'\mathbf{x}_s^{(k)}). \tag{3.18}$$

For comparison, other distributions such as lognormal, or Burr could also be used to model severity.

3.5 Incurred But Not Reported (IBNR)

So far, we have fitted the reporting delay, claim count and claim severity which is sufficient for the projection of INBR losses. It is enough for us to get the IBNR losses for risk class k:

$$\widehat{\mathbf{IBNR}}^{(k)} = \sum_{i=1}^{m} \sum_{j=m+1-i}^{m-1} \hat{N}_{i,j}^{(k)}\ \hat{\mu}^{(k)}. \tag{3.19}$$

Summing up all the K equations in the whole portfolio, we get:

$$\widehat{\mathbf{IBNR}} = \sum_{k=1}^{K} \widehat{\mathbf{IBNR}}^{(k)} = \sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{j=m+1-i}^{m-1} \hat{N}_{i,j}^{(k)}\ \hat{\mu}^{(k)}. \tag{3.20}$$

3.6 Settlement Delay

The settlement delay (SD) is defined as the relative year to the notification of a claim in which the claim was closed, starting from 0. We use the same method to model the settlement delay as for the reporting delay in Section 3.1. Finally we will get the estimates of

$$q_j^{(k)}, \qquad k = 1, 2, \dots, K, j = 0, 1, \dots, d, \tag{3.21}$$

where $q_j^{(k)}$ denotes the probability for the claim in the risk class k to have the settlement delay equal to j.

3.7 Reported But Not Settled (RBNS)

With the estimation of settlement delay probabilities, we can give the IBNR and RBNS entries in the reserve matrix for the overall business.

The IBNR entries for the reserve matrix are given by:

$$\hat{Z}_{i,j}^{ibnr} = \sum_{k=1}^{K} \sum_{l=0}^{i-m+j-1} \hat{N}_{i,j-l}^{(k)} \hat{\mu}^{(k)} \hat{q}_l^{(k)}, \qquad \text{for } i + j > m, \tag{3.22}$$

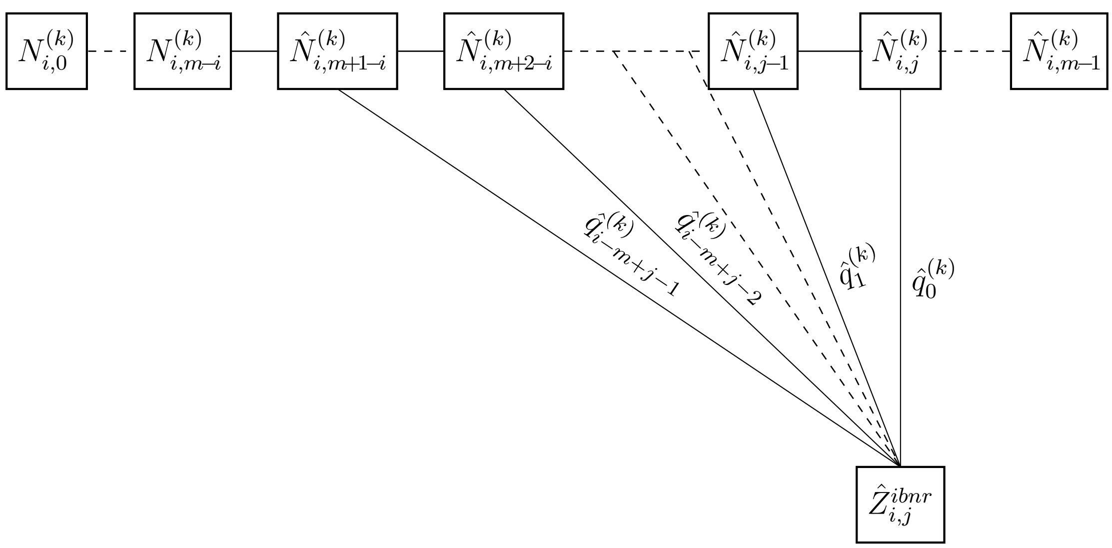

Figure 3.1: IBNR Entries

and the RBNS entries for the reserve matrix are given by:

$$\hat{Z}_{i,j}^{rbns} = \sum_{k=1}^{K} \sum_{l=i-m+j}^{j} N_{i,j-l}^{(k)} \hat{\mu}^{(k)} \hat{q}_l^{(k)}, \qquad \text{for } i + j > m, \tag{3.23}$$

here $\hat{q}_l^{(k)}$ is defined as 0 if $l > d$.

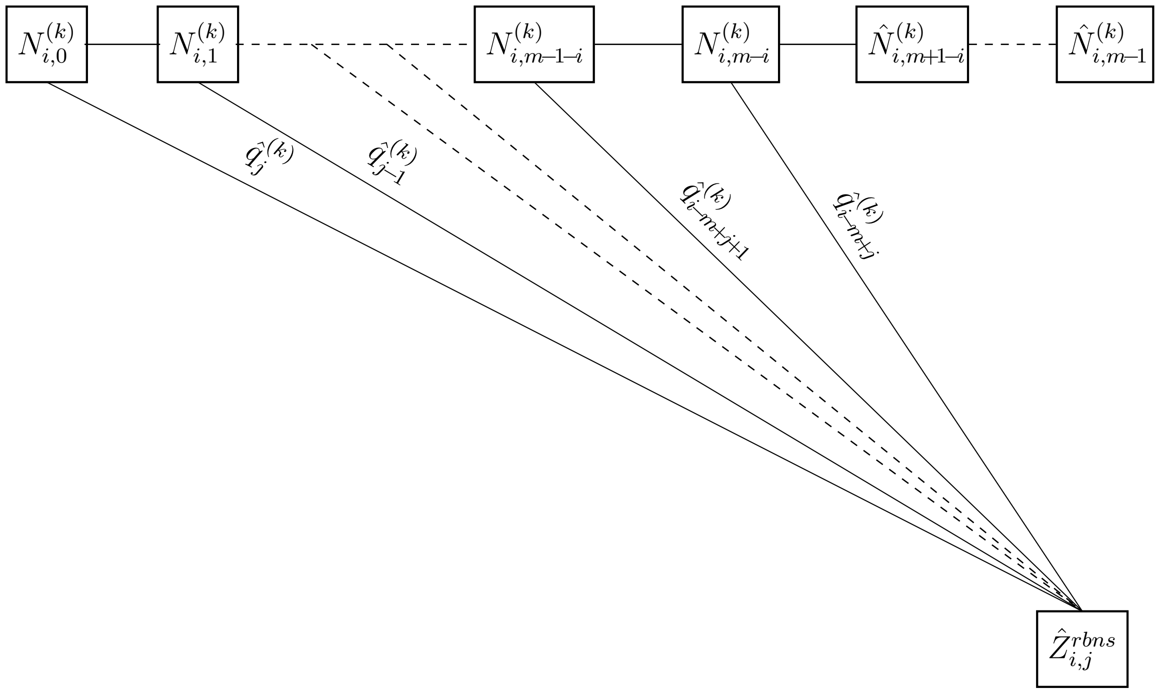

Figure 3.2: RBNS Entries

Finally, we can give the estimations of each entry for the lower-right half of the reserve matrix:

$$\hat{Z}_{i,j} = \hat{Z}_{i,j}^{ibnr} + \hat{Z}_{i,j}^{rbns}, \qquad \text{for } i + j > m. \tag{3.24}$$

Chapter 4

Simulation Results

In this chapter, we use a Monte Carlo simulation method to compare different loss reserve estimation methods, including the traditional chain ladder method (CLM), the double chain ladder (DCL) by Miranda et al. (2013), the aggregate Tweedie GLMs reserve method and our individual GLMs reserve method. Since the properties of the estimators cannot be studied analytically, statistical simulation is a well-accepted technique for comparing various methods of estimation. Our approach is similar to those of Stanard (1985) and Narayan and Warthen (1997).

When we are estimating the reserves, only the top-left half of the loss triangle is available to us as data, and is used to estimate the lower-right half of the triangle, which represents the projection of ultimate losses.

We compute the deviations of the estimated reserves from the empirical reserves. Finally, we use several criteria to compare the deviations of estimated versus empirical under various reserving methods.

4.1 Simulation of Random Loss Triangles

Our loss triangles are simulated based on the ideas of Narayan and Warthen (1997). Our method is described below:

1. Generate $N_i^{(k)}$, the frequency of claims for risk class k and accident year i as a Poisson variate with mean $w_i^{(k)} \exp(\boldsymbol{\beta}'_f \mathbf{x}_f^{(k)})$, where $w_i^{(k)}$ is the number of insured

for risk class k and accident year i.

2. For each claim, generate Weibull distributed variates RD and SD for the reporting delay and settlement delay.

3. The claim amounts for a claim follow a gamma distribution with mean $\exp(\boldsymbol{\beta}_s'\mathbf{x}_s^{(k)})$ and scale parameter v.

4. Repeat the steps through for all the risk classes and accident years.

5. These claims are added up to create the aggregate reported claim counts matrix and aggregate paid losses matrix.

See Appendix E for the detailed descriptions of parameters in this simulation method.

4.2 Comparison of Methods

We generate 1,000 realizations of hypothetical reserve data for this simulation method. For each of the 1,000 sets of hypothetical data, the reserves were estimated by CLM, DCL, aggregate Tweedie GLM, and our individual GLMs. The deviations of the reserve estimates using different methods from the actual reserves are computed.

Then we give a table of the basic descriptive statistics for these deviations:

	Individual GLMs	CLM	DCL	Tweedie GLM
Min.	-64,613.0	-97,128.0	-89,260.0	-97,097.0
1st Qu.	-17,767.0	-24,095.5	-22,062.8	-24,171.5
Median	-942.5	-459.5	765.5	-639.0
Mean	-774.5	-419.8	1,378.6	-447.6
3rd Qu.	14,820.8	22,215.5	22,803.2	22,692.2
Max.	71,787.0	114,752.0	113,492.0	114,896.0
Variance	5.563E8	1.225E9	1.119E9	1.225E9

Table 4.1: Summary of Deviations

We continue to produce a box plots of the calculated deviations:

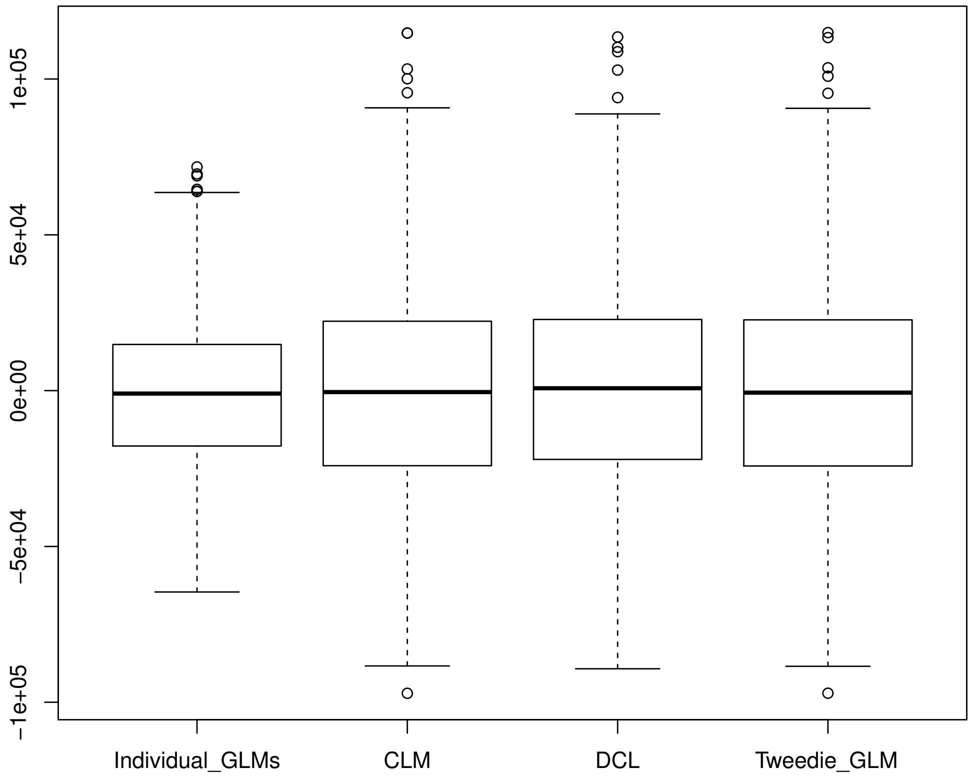

Figure 4.1: Box Plot of Deviations

A box plot is a convenient way of graphically describing groups of numerical data through their quartiles. The bottom and top of the box are the first and third quartiles, and the band inside the box is the median. The ends of the whiskers represent the lowest datum still within 1.5 Inter Quartile Range (IQR) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile.

The histograms of these 1000 deviations for each method are plotted below:

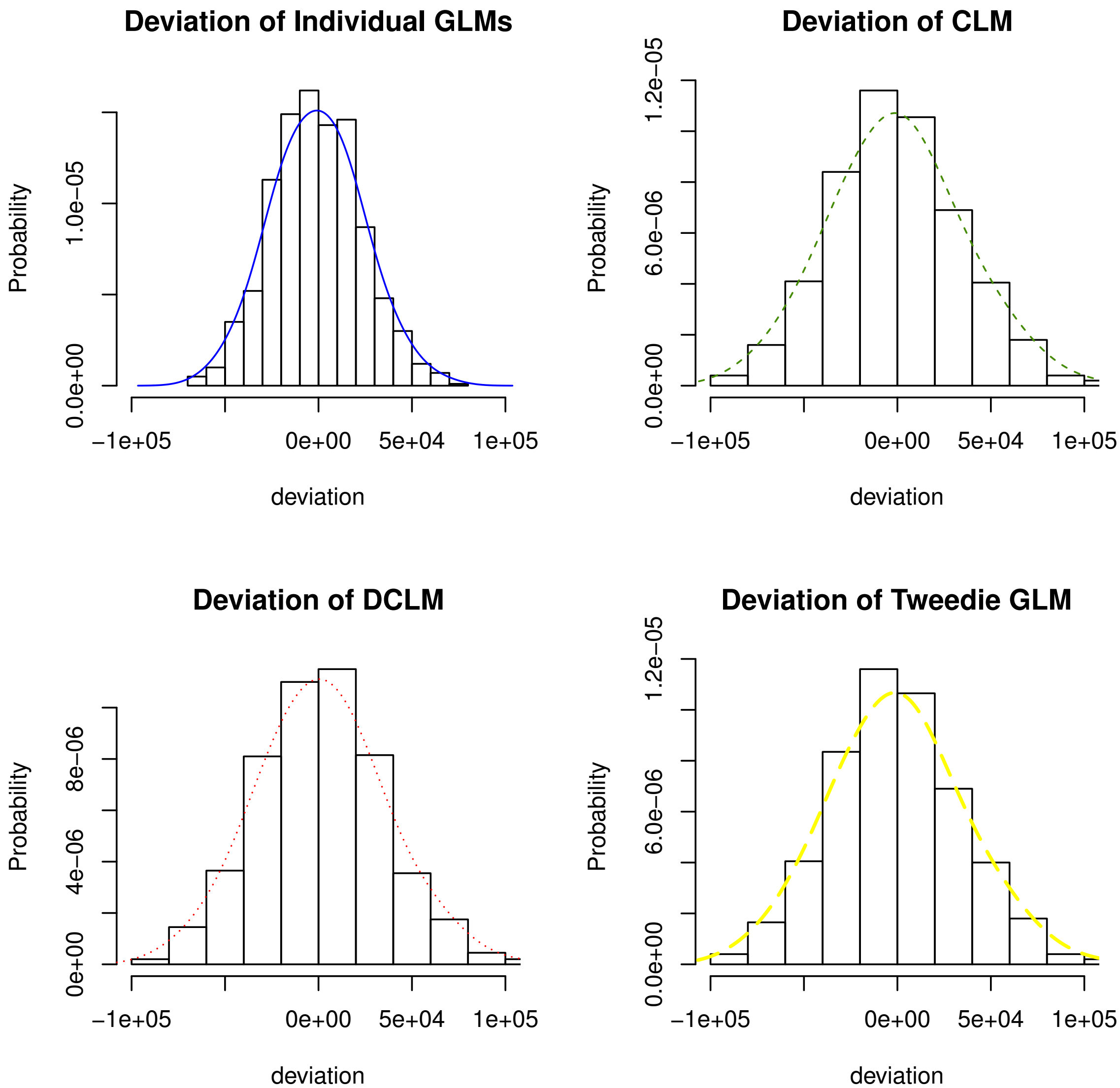

Figure 4.2: Histograms of Deviations

Then we fit a density curve to each histogram using smooth kernel estimates, and collect all the density curves in one picture:

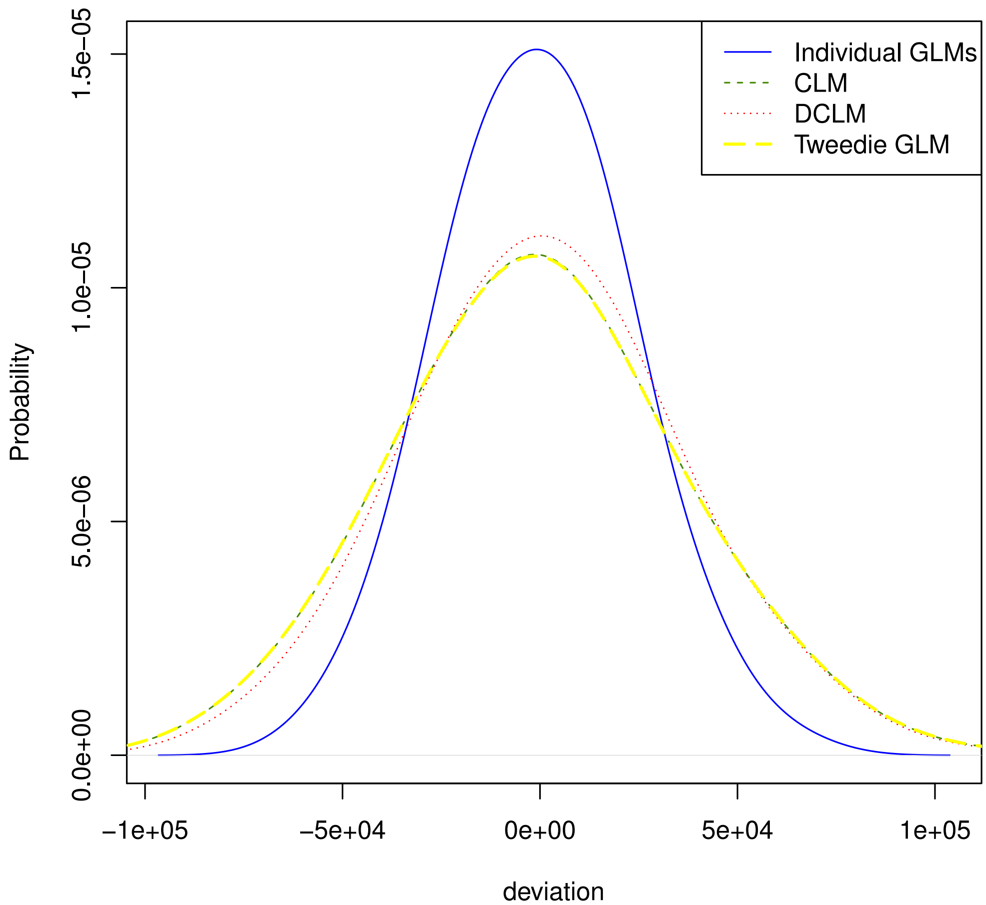

Figure 4.3: Kernel Density Plot of Deviation

From the statistics and plots above, we see that:

- The density function of our individual GLMs deviations are peaked, concentrated around 0 and have lighter tails. The word "lighter" is different from light (heavy) tailed in probability theory. It only means to be lighter than the traditional reserve methods.

- The deviations for the CLM and Tweedie aggregate reserve methods are very close. The reason is that CLM is nothing but an aggregate Poisson GLM reserve method, and the Poisson GLM is a member of Tweedie GLMs. For our simulation parameter choices the differences between CLM and Tweedie aggregate reserve method are not significant enough to be noticed.

- The DCL method offers a better performance than CLM, but it is hard for it to compete with our individual GLMs reserve method.

4.3 Changing Exposures

Next we let the weight $w_i^{(k)}$ change with the accident year i, i.e. the weights of different risk groups in the overall business change with time. A lognormal distributed multiplicative factor is multiplied to the base weight $w_0^{(k)}$ to get the real weight for each accident year:

$$w_i^{(k)} = w_0^{(k)} \ \ln N(-0.045, 0.3), \qquad i = 1, 2, \ldots, m, \ k = 1, 2, \ldots, K.$$

We compare the fitted density curves of the reserves in the most recent accident year with fixed exposures and changing exposures for each method:

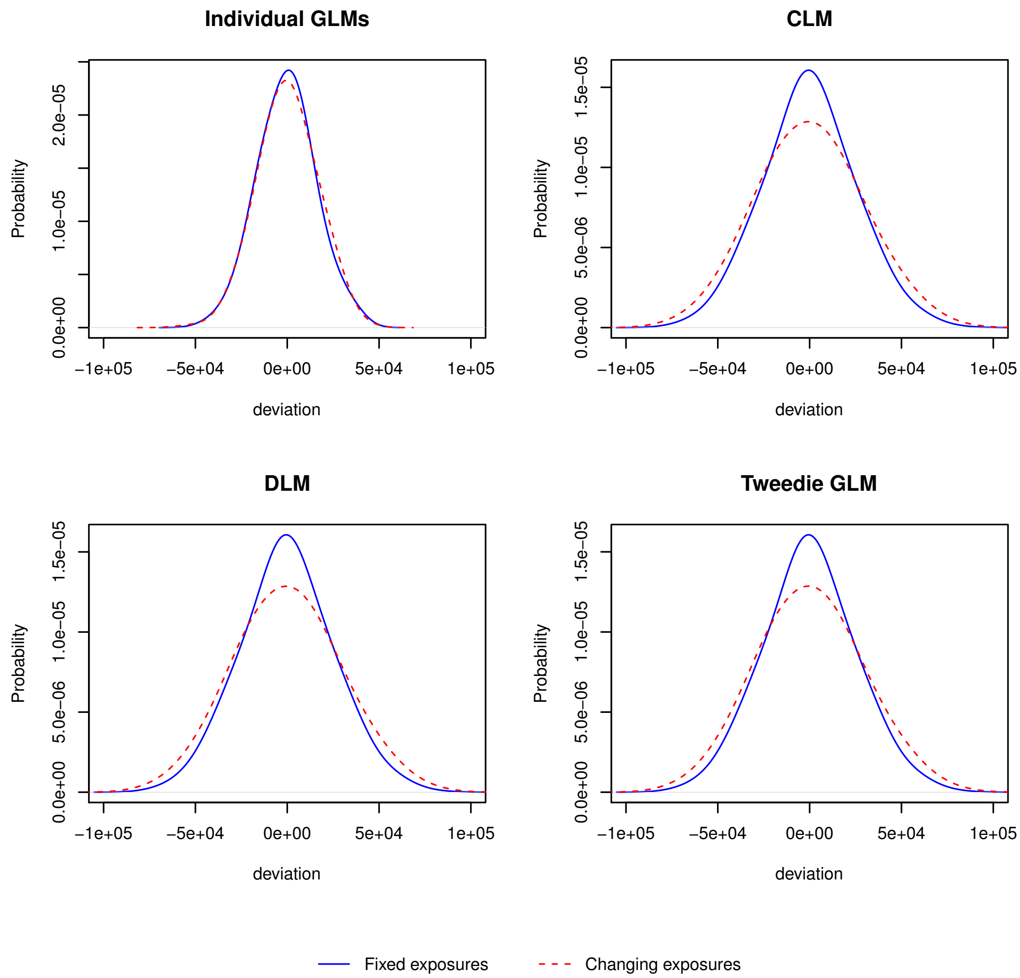

Figure 4.4: Comparison of Changing Exposures and Fixed Exposures

The blue lines represent the deviations when the exposures are fixed and the red line represent the deviations when the exposures are changing with time. From the above pictures, we see that our individual GLMs reserve method can deal with the changing exposures, but the traditional methods do not work properly.

4.4 Excess Zeros for Claim Number

In this section, we make a minor change to the simulation method described in Section 4.1. Rather than using Poisson variates for claim counts, we use zero inflated Poisson variates for claim counts.

In the first scenario, we let the probability of extra zeros $\pi_0 = 0.2$. We give the histograms below:

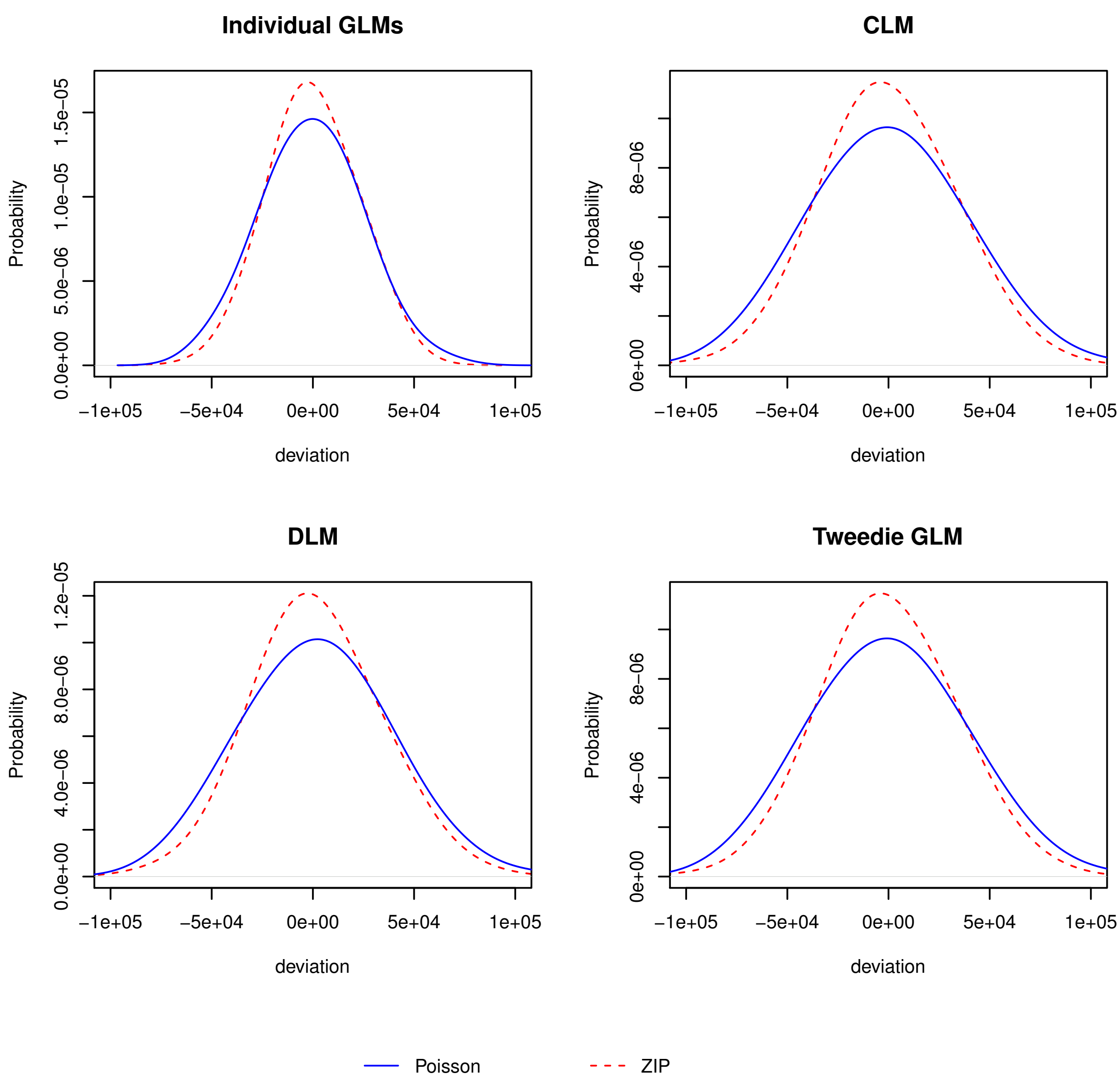

Figure 4.5: Densities of Poisson and Zero Inflated Poisson ($\pi_0 = 0.2$)

The blue lines represent the deviations when a Poisson assumption is used for claim counts and the red line represent the deviations when the Zero Inflated Poisson assumption is used. We also calculate the variances of the new deviations and their ratios to the original variances:

Method	Individual GLMs	CLM	DCL	Tweedie GLM
VAR (Poisson)	5.563E8	1.225E9	1.119E9	1.225E9
VAR (ZIP)	4.327E8	9.420E8	8.885E8	9.423E8
Ratio	0.778	0.770	0.794	0.770

Table 4.2: Variances with Poisson and Zero Inflated Poisson ($\pi_0 = 0.2$)

Then we do the same for $\pi = 0.6$:

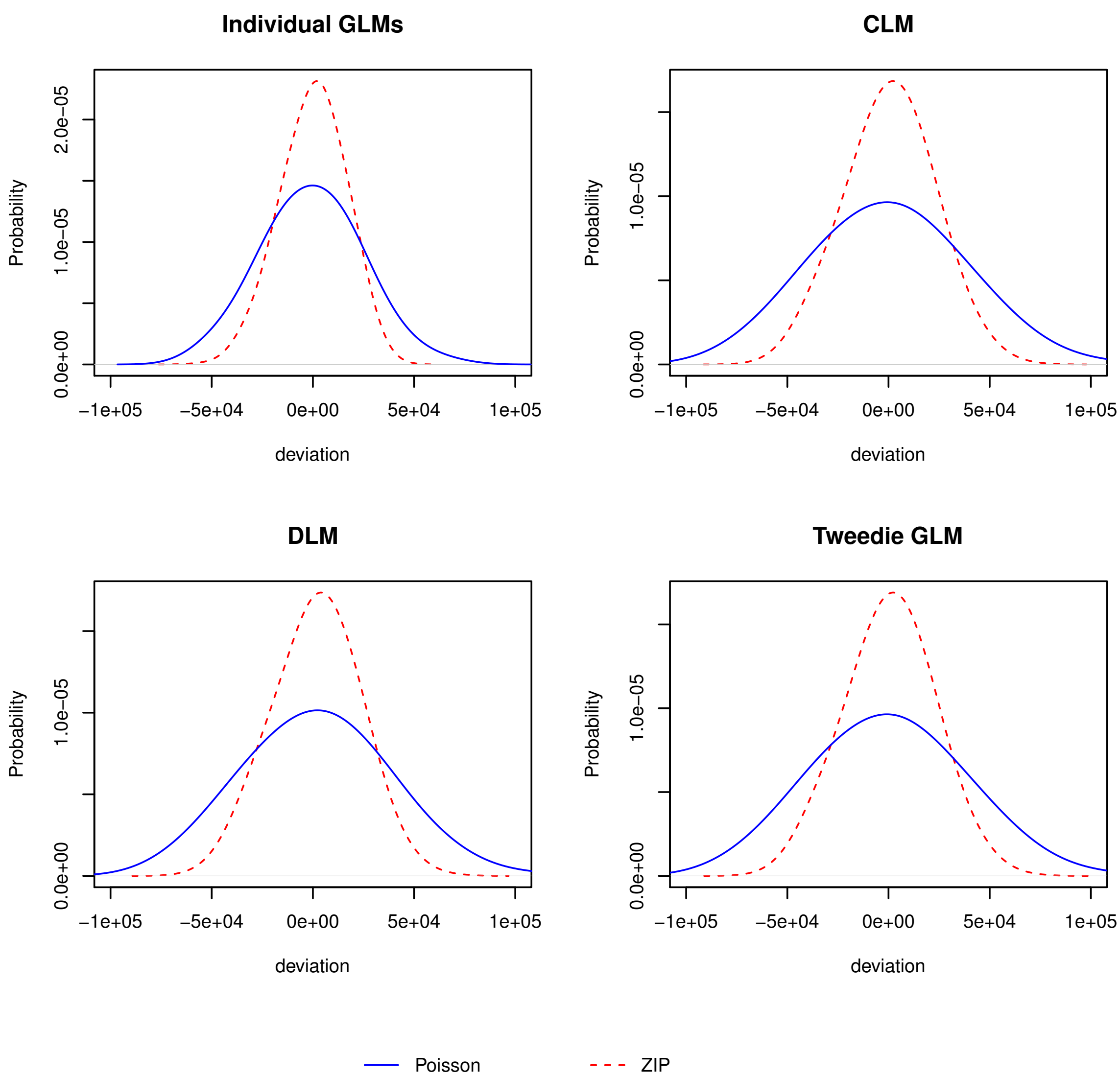

Figure 4.6: Densities of Poisson and Zero Inflated Poisson ($\pi_0 = 0.6$)

Method	Individual GLMs	CLM	DCL	Tweedie GLM
VAR (Poisson)	5.563E8	1.225E9	1.119E9	1.225E9
VAR (ZIP)	2.212E8	4.683E8	4.410E8	4.669E8
Ratio	0.398	0.382	0.394	0.381

Table 4.3: Variances with Poisson and Zero Inflated Poisson ($\pi_0 = 0.6$)

From these two examples, we see that:

- With excess zeros, the variances of deviations for all the four methods decrease simultaneously.
- The variances of deviations for all the four methods shrink at almost the same rate.
- Since the original variance of the deviations using our individual GLMs reserve method is smaller than the other three, after we add the excess zeros assumption to the claim counts, it still offers a better performance then the other methods.
- The rates at which the variances shrink are approximately equal to $1 - \pi_0$.

4.5 Conclusions of Simulation

Our individual GLMs reserve method provides an useful tool to structurally estimate the loss reserve and also a stable and efficient way to improve the estimates of ultimate loss reserve in actual applications.

Our model leverages the frequency and severity estimation, both in ratemaking and loss reserving, making it more consistent and easier to interpret.

A reserve estimation method that incorporates some underlying assumptions about the claim process will provide a better estimate of the loss reserve if those assumptions are satisfied.

It is not suggested that actuaries blindly apply the chain ladder method to aggregate data. Efficiency is lost by the throwing away of a great variety of detailed information on the actual claim process. The estimates of development factors in CLM are vulnerable to outliers. The bias in one development factor will be passed to the projections of losses in the subsequent development periods.

When the exposures of different risk groups are changing with accident years, the weaknesses of traditional chain ladder methods are exacerbated.

Finally we really hope that the individual GLMs reserve method in the actuarial field will become popular and used in the future.

References

Andersen, E. B. (1970) "Asymptotic properties of conditional maximum-likelihood estimators." Journal of the Royal Statistical Society Series B (Methodological), **vol. 32(2)**: pp. 283–301.

Antonio, K. and Beirlant, J. (2007) "Actuarial statistics with generalized linear mixed models." Insurance: Mathematics and Economics, **vol. 40(1)**: pp. 58–76.

Antonio, K. and Plat, R. (2014) "Micro-level stochastic loss reserving for general insurance." Scandinavian Actuarial Journal, **vol. 2014(7)**: pp. 649–669.

Booth, J. G. and Hobert, J. P. (1998) "Standard errors of prediction in generalized linear mixed models." Journal of the American Statistical Association, **vol. 93(441)**: pp. 262–272.

Bornhuetter, R. and Ferguson, R. E. (1972) "The actuary and IBNR." Proceedings of the Casualty Actuarial Society, **vol. 59**: pp. 181–195.

Boucher, J.-P., Denuit, M. and Guillén, M. (2007) "Risk classification for claim counts: a comparative analysis of various zero-inflated mixed Poisson and hurdle models." North American Actuarial Journal, **vol. 11(4)**: pp. 110–131.

Breslow, N. and Lin, X. (1995) "Bias correction in generalised linear mixed models with a single component of dispersion." Biometrika, **vol. 82(1)**: pp. 81–91.

Bühlmann, H. (1967) "Experience rating and credibility." ASTIN Bulletin, **vol. 4(3)**: pp. 199–207.

De Vylder, F. (1981) “Practical credibility theory with emphasis on optimal parameter estimation.” ASTIN Bulletin, **vol. 12**: pp. 115–131.

Demindenko, E. (2004) Mixed Models: Theory and Applications. John Wiley & Sons, Hobeken, New Jersey.

Fahrmeir, L. and Kaufmann, H. (1985) “Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models.” Annals of Statistics, **vol. 13(1)**: pp. 342–368.

Frees, E. W., Young, V. R. and Luo, Y. (1999) “A longitudinal data analysis interpretation of credibility models.” Insurance: Mathematics and Economics, **vol. 24(3)**: pp. 229–247.

Fuchs, S. (2014) “Consistent loss prediction for a portfolio and its subportfolios.” Scandinavian Actuarial Journal, **(6)**: pp. 561–581.

Gigante, P., Picech, L. and Sigalotti, L. (2013) “Prediction error for credible claims reserves: an h-likelihood approach.” European Actuarial Journal, **vol. 3(2)**: pp. 453–470.

Guszcza, J. and Lommele, J. (2006) “Loss reserving using claim-level data.” Casualty Actuarial Society Forum, Fall 2006: pp. 111–140.

Hachemeister, C. (1975) “Credibility for regression models with application to trend.” In “Credibility, theory and applications, Proceedings of the Berkeley Actuarial Research Conference on Credibility,” pp. 129–163.

Jewell, W. S. (1974) “Credible means are exact Bayesian for exponential families.” ASTIN Bulletin, **vol. 8(1)**: pp. 77–90.

Jiang, J. (2007) Linear and Generalized Linear Mixed Models and their Applications. Springer, New York.

Jorgensen, B. (1987) “Exponential dispersion models.” Journal of the Royal Statistical Society Series B (Methodological): pp. 127–162.

Klinker, F. (2010) "Generalized linear mixed models for ratemaking: a means of introducing credibility into a generalized linear model setting." Casualty Actuarial Society E-Forum, Winter 2011, **vol. 2**: pp. 1–25.

Lancaster, T. (2000) "The incidental parameter problem since 1948." Journal of Econometrics, **vol. 95(2)**: pp. 391–413.

Lin, X. and Breslow, N. (1996) "Bias correction in generalized linear mixed models with multiple components of dispersion." Journal of the American Statistical Association, **vol. 91(435)**: pp. 1007–1016.

Liu, Q. and Pierce, D. (1994) "A note on Gauss-Hermite quadrature." Biometrika, **vol. 81(3)**: pp. 624–629.

McCullogh, C. and Searle, S. (2001) Generalized Linear and Mixed Models. Wiley, New York.

Merz, M. and Wüthrich, M. (2010) "Paid-incurred chain claims reserving method." Insurance: Mathematics and Economics, **vol. 46(3)**: pp. 568–579.

Miranda, M. D. M., Nielsen, B., Nielsen, J. P. and Verrall, R. (2011) "Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers." ASTIN Bulletin, **vol. 41(1)**: pp. 107–129.

Miranda, M. D. M., Nielsen, J. and Verrall, R. (2012) "Double chain ladder." ASTIN Bulletin, **vol. 42(1)**: pp. 59–76.

Miranda, M. D. M., Nielsen, J. P. and Verrall, R. (2013) "Double chain ladder and Bornhuetter-Ferguson." North American Actuarial Journal, **vol. 17(2)**: pp. 101–113.

Mowbray, A. (1914) "How extensive a payroll exposure is necessary to give a dependable pure premium." Proceedings of the Casualty Actuarial Society, **vol. 1**: pp. 24–30.

Narayan, P. and Warthen, T. V. (1997) “A comparative study of the performance of loss reserving methods through simulation.” Casualty Actuarial Society Forum, Summer 1997, **vol. 1**: pp. 175–196.

Nelder, J. and Verrall, R. (1997) “Credibility theory and genralized linear models.” ASTIN Bulletin, **vol. 27(1)**: pp. 71–82.

Ohlsson, E. (2008) “Combining generalized linear models and credibility models in practice.” Scandinavian Actuarial Journal, **vol. 2008(4)**: pp. 301–314.

Ohlsson, E. and Johansson, B. (2006) “Exact credibility and Tweedie models.” ASTIN Bulletin, **vol. 36(1)**: pp. 121–133.

Owen, D. (1980) “A table of normal integrals: A table.” Communications in Statistics-Simulation and Computation, **vol. 9(4)**: pp. 389–419.

Pigeon, M., Antonio, K. and Denuit, M. (2013) “Individual loss reserving with the multivariate skew normal framework.” ASTIN Bulletin, **vol. 43(3)**: pp. 399–428.

Quarg, G. and Mack, T. (2008) “Munich chain ladder: a reserving method that reduces the gap between INBR projections based on paid losses and IBNR projections based on incurred losses.” Variance, **vol. 2(2)**: pp. 266–299.

Renshaw, A. and Verrall, R. (1998) “A stochastic model underlying the chain ladder technique.” British Actuarial Journal, **vol. 4**: pp. 903–923.

Rosenberg, P. (1990) “A simple correction of AIDS surveillance data for reporting delays.” Journal of Acquired Immune Deficiency Syndromes, **vol. 3**: pp. 49–54.

Sagae, M. and Tanabe, K. (1994) “Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices.” Linear and Multilinear Algebra, **vol. 37(4)**: pp. 279–282.

SAS/ETS (2010a) COUNTREG Procedure. SAS Institute Inc., Cary, NC, 9.22 ed. `http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_countreg_sect001.htm`.

SAS/ETS (2010b) SEVERITY Procedure. SAS Institute Inc., Cary, NC, 9.22 ed. http://support.sas.com/documentation/cdl/en/etsug/63348/PDF/default/etsug.pdf.

SAS/IML (2010) Interactive Matrix Language. SAS Institute Inc., Cary, NC, 9.22 ed. http://support.sas.com/documentation/cdl/en/imlug/63541/PDF/default/imlug.pdf.

SAS/STAT (2006) GLIMMIX Procedure. SAS Institute Inc., Cary, NC, 9.1 ed. http://support.sas.com/rnd/app/papers/glimmix.pdf.

SAS/STAT (2009) SAS/STAT 9.2 User Guide. 2 ed. SAS Institute Inc., Cary, NC. http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm.

SAS/STAT (2013) GENMOD Procedure. SAS Institute Inc., Cary, NC, 13.1 ed. http://support.sas.com/documentation/onlinedoc/stat/131/genmod.pdf.

Stanard, J. N. (1985) A simulation test of prediction errors of loss reserve estimation techniques. Ph.D. thesis, New York University, Graduate School of Business Administration.

Verdonck, T., Van Wouwe, M. and Dhaene, J. (2009) “A robustification of the chain-ladder method.” North American Actuarial Journal, **vol. 13(2)**: pp. 280–298.

Vonesh, E. (1996) “A note on the use of Laplace’s approximation for nonlinear mixed-effects models.” Biometrika, **vol. 83(2)**: pp. 447–452.

Wolfinger, R. and O’Connell, M. (1993) “Generalized linear mixed models: a pseudo-likelihood approach.” Journal of Statistical Computation and Simulation, **vol. 4**: pp. 233–243.

Wüthrich, M. (2003) “Claim reserving using Tweedie’s compound Poisson model.” ASTIN Bulletin, **vol. 33(2)**: pp. 331–346.

Zeger, S. and Karim, M. (1991) "Generalized linear models with random effects; a Gibbs sampling approach." Journal of the American Statistical Association, **vol. 86(413)**: pp. 79–86.

Zhou, J. and Garrido, J. (2009a) "Full credibility with generalized linear and mixed models." ASTIN Bulletin, **vol. 39(1)**: pp. 61–80.

Zhou, J. and Garrido, J. (2009b) "A loss reserving method based on generalized linear models." Tech. Rep. 2, Concordia University. `http://www.mathstat.concordia.ca/resources-and-links/technical-reports/REPORTS/2_09.pdf`.

Appendix A

Derivation of Conditional Variance

For the approximation of $\pi_{ij}(\mathbf{u}_i)$ in (2.31), we first decompose the difference $\hat{\eta}_{ij} - \eta_{ij}$ into three parts:

$$\hat{\eta}_{ij} - \eta_{ij} = \left(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} - \mathbf{x}'_{ij}\boldsymbol{\beta}\right) + \left(\mathbf{z}'_{ij}\hat{\hat{\mathbf{u}}}_i - \mathbf{z}'_{ij}\hat{\mathbf{u}}_i\right) + \left(\mathbf{z}'_{ij}\hat{\mathbf{u}}_i - \mathbf{z}'_{ij}\mathbf{u}_i\right). \tag{A.1}$$

The first term arises from the difference between the ML estimator $\hat{\boldsymbol{\beta}}$ and the true fixed effect coefficient $\boldsymbol{\beta}$. The third error term represents the difference between the random effect estimator $\hat{\mathbf{u}}_i$ and true random effect $\mathbf{u}_i$. Finally, the second error term is a reflection of the difference between plugging the estimator $\hat{\boldsymbol{\beta}}$ in (2.28) or using the true $\boldsymbol{\beta}$ (we find σ is not used here). However, the second term is not usually noticed.

Let $\ell(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0|\mathbf{y}_i) = \sum_{j=1}^{n_i} \log\left\{f(y_{ij}|\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0^2)\right\}$. It is a log-likelihood function with arguments $\mathbf{u}_i$, $\boldsymbol{\beta}$ and σ_0. We know that:

$$\frac{\partial}{\partial \mathbf{u}_i}\ell(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0|\mathbf{y}_i) = \sum_{j=1}^{n_i} \frac{w_{ij}}{\sigma_0^2}\frac{y_{ij}-\mu_{ij}}{V(\mu_{ij})}\frac{1}{g'(\mu_{ij})}\mathbf{z}_{ij}, \tag{A.2}$$

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i.$$

Notice here that $\hat{\mathbf{u}}_i$ is the solution to $\dfrac{\partial}{\partial \mathbf{u}_i}\ell(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0|\mathbf{y}_i) = \mathbf{0}$ based on true $\boldsymbol{\beta}$ and σ_0:

$$\frac{\partial}{\partial \mathbf{u}_i}\ell\big(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0\big|\mathbf{y}_i\big)\Big|_{\mathbf{u}_i=\hat{\mathbf{u}}_i} = \mathbf{0},$$

and $\hat{\hat{\mathbf{u}}}_i$ is the solution based on the MLE $\hat{\boldsymbol{\psi}}$:

$$\frac{\partial}{\partial \mathbf{u}_i}\ell\big(\mathbf{u}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_0\big|\mathbf{y}_i\big)\Big|_{\mathbf{u}_i=\hat{\hat{\mathbf{u}}}_i} = \mathbf{0},$$

here σ_0 and $\hat{\sigma}_0$ are scale parameters so they are not used.

Lemma A.1. *Applying the inverse function theorem, we have that:*

$$\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i$$
$$\approx \Big[\frac{\partial^2}{\partial \mathbf{u}_i \partial \mathbf{u}_i'} \ell\big(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0 | \mathbf{y}_i\big)\Big| \mathbf{u}_i = \hat{\mathbf{u}}_i\Big]^{-1} \Big[-\frac{\partial^2}{\partial \mathbf{u}_i \partial \boldsymbol{\beta}_i'} \ell\big(\mathbf{u}_i, \boldsymbol{\beta}, \sigma_0 | \mathbf{y}_i\big)\Big| \mathbf{u}_i = \hat{\mathbf{u}}_i\Big](\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{A.3}$$

where the approximation error is $o(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

From further calculations we get that:

$$\frac{\partial^2}{\partial \mathbf{u}_i \partial \mathbf{u}_i'} \ell(\hat{\mathbf{u}}_i, \boldsymbol{\beta}, \sigma_0 | \mathbf{y}_i) = -\mathbf{Z}_i' \mathbf{W_1} \mathbf{Z}_i, \qquad \mathbf{Z}_i = (z_{i1}, z_{i2}, \ldots, z_{in_i})',$$
$$\frac{\partial^2}{\partial \mathbf{u}_i \partial \boldsymbol{\beta}_i'} (\hat{\mathbf{u}}_i, \boldsymbol{\beta}, \sigma_0 | \mathbf{y}_i) = -\mathbf{Z}_i' \mathbf{W_1} \mathbf{X}_i, \qquad \mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})',$$

and $\mathbf{W_1} = diag(w_1^1, w_2^1, \ldots, w_{n_i}^1)$ with

$$w_j^1 = \frac{w_{ij}}{\sigma_0^2 \, V(\mu_{ij}^1)(g'(\mu_{ij}^1))^2} + w_{ij}(y_{ij} - \mu_{ij}^1) \frac{V(\mu_{ij}^1) g''(\mu_{ij}^1) + V'(\mu_{ij}^1) g'(\mu_{ij}^1)}{\sigma_0^2 \, (V(\mu_{ij}^1))^2 (g'(\mu_{ij}^1))^3},$$
$$\mu_{ij}^1 = g^{-1}(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \hat{\mathbf{u}}_i).$$

Denote by

$$\mathbf{F_1} = \mathbf{Z}_i' \mathbf{W_1} \mathbf{Z}_i, \qquad \mathbf{F_2} = \mathbf{Z}_i' \mathbf{W_2} \mathbf{Z}_i \qquad \text{and} \qquad \mathbf{F_0} = \mathbf{F} = \mathbf{Z}_i' \mathbf{W_0} \mathbf{Z}_i,$$

where $\mathbf{W_2} = diag(w_1^2, w_2^2, ..., w_{n_i}^2)$, with

$$w_j^2 = \frac{w_{ij}}{\sigma_0^2 \, V(\mu_{ij})(g'(\mu_{ij}))^2} + w_{ij}(y_i - \mu_{ij}) \frac{V(\mu_{ij}) g''(\mu_{ij}) + V'(\mu_{ij}) g'(\mu_{ij})}{\sigma_0^2 \, (V(\mu_{ij}))^2 (g'(\mu_{ij}))^3},$$
$$\mu_{ij} = g^{-1}(\mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{u}_i),$$

and $\mathbf{W_0} = diag(w_1^0, w_2^0, ..., w_{n_i}^0)$, with $w_j^0 = \dfrac{w_{ij}}{\sigma_0^2 \, V(\mu_{ij})(g'(\mu_{ij}))^2}$.
$\mathbf{F}$ is the information matrix in the GLM fitting process in (2.28) and let $\mathbf{F}^{\frac{1}{2}}$ ($\mathbf{F}^{\frac{1}{2}'}$) be its left (respective right) square root.

Fahrmeir and Kaufmann (1985) studied the consistency and asymptotic normality of maximum likelihood estimator in generalized linear models. They showed the following results.

Lemma A.2. *Under some regularity conditions:*

1. $\mathbf{F}^{\frac{1}{2}'}(\hat{\mathbf{u}}_i - \mathbf{u}_i) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{I}$ *is the identity matrix.*

2. $\mathbb{E}(\mathbf{F_2}) = \mathbf{F}$, *and* $\mathbf{F_2}/n_i$ *converges to* $\mathbf{F}/n_i$ *for large* n_i.

Using the first property, we could regard the asymptotic behavior of $\hat{\mathbf{u}}_i - \mathbf{u}_i$ as $\mathbf{N}(\mathbf{0}, \mathbf{F}^{-1})$, see Zhou and Garrido (2009a), and $\mathbf{F}^{-1}$ converges to $\mathbf{0}$ at a speed of $O(\frac{1}{n_i})$. This makes it possible to replace $\mathbf{W_1}$ in (A.3) by $\mathbf{W_2}$. We continue to replace $\mathbf{W_2}$ by $\mathbf{W_0}$ according to the second property. Finally we have:

Theorem A.1. *Under some regularity conditions,*

$$\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i \approx (-\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{W_0}\mathbf{X}_i)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{A.4}$$

The approximation error is

$$o(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + O_p(n_i^{-\frac{1}{2}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Proposition A.1. *We list some special cases of formula (A.4):*

1. *When* $\mathbf{X}_i = \mathbf{Z}_i$, *formula (A.4) gives* $\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i \approx -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, *but obviously they are equal. Since in this special case* $\boldsymbol{\beta} + \hat{\mathbf{u}}_i$ *enter into the regressor as a unit, no matter if the real* $\boldsymbol{\beta}$ *or its ML estimator* $\hat{\boldsymbol{\beta}}$ *is used, the sum* $\boldsymbol{\beta} + \hat{\mathbf{u}}_i$ *is not changed, thus equal to* $\hat{\boldsymbol{\beta}} + \hat{\hat{\mathbf{u}}}_i$.

2. *When* $\mathbf{X}_i$ *is part of the columns of* $\mathbf{Z}_i$, *for example:*

$$\mathbf{X}_i = \mathbf{Z}_i \begin{pmatrix} \mathbf{I}_p \\ \mathbf{0}_{(q-p)\times p} \end{pmatrix},$$

then formula (A.4) turns into

$$\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i \approx - \begin{pmatrix} \mathbf{I}_p \\ \mathbf{0}_{(q-p)\times p} \end{pmatrix} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

However, for the same reason we know they are actually equal.

3. *If the exponential dispersion distribution is normal and the link function is canonical, then* $w_j^0 = \frac{w_{ij}}{\sigma_0^2}$ *and the model is:*

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u} + \boldsymbol{\epsilon}_i, \qquad \epsilon_{ij} \sim N(0, \sigma_0^2/w_{ij}).$$

It is easy to see that the MLEs are:

$$\hat{\mathbf{u}}_i = (\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\ \mathbf{W_0}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}),$$
$$\hat{\hat{\mathbf{u}}}_i = (\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\ \mathbf{W_0}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}).$$

Subtracting these two equations, we will get the same result:

$$\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i = (-\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{W_0}\mathbf{X}_i)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{A.5}$$

Note here that given $\mathbf{u}_i$ in (A.4) the term $(-\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{W_0}\mathbf{X}_i)$ is constant in value, as it does not depend on $\mathbf{y}_i$ any more. It helps us easily derive the asymptotic variance of $\hat{\hat{\mathbf{u}}}_i - \hat{\mathbf{u}}_i$ as well as makes $(\mathbf{x}_{ij}'\hat{\boldsymbol{\beta}} - \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\hat{\hat{\mathbf{u}}}_i - \mathbf{z}_{ij}'\hat{\mathbf{u}}_i)$ independent of $\hat{\mathbf{u}}_i - \mathbf{u}_i$ conditional on $\mathbf{u}_i$ ($\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ depends on $\mathbf{y}_{-i}$, $\hat{\mathbf{u}}_i - \mathbf{u}_i$ depends on $\mathbf{y}_i$).

Combining (A.4) with the first term in (A.1), we have:

$$\begin{aligned}&\left(\mathbf{x}_{ij}'\hat{\boldsymbol{\beta}} - \mathbf{x}_{ij}'\boldsymbol{\beta}\right) + \left(\mathbf{z}_{ij}'\hat{\hat{\mathbf{u}}}_i - \mathbf{z}_{ij}'\hat{\mathbf{u}}_i\right)\\ \approx&\big\{\mathbf{x}_{ij}' + \mathbf{z}_{ij}'(-\mathbf{Z}_i'\mathbf{W}_0\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{W}_0\mathbf{X}_i)\big\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).\end{aligned} \tag{A.6}$$

If we denote by $\mathbf{A}_{ij}' = \mathbf{x}_{ij}' + \mathbf{z}_{ij}'(-\mathbf{Z}_i'\mathbf{W_0}\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{W_0}\mathbf{X}_i)$, then

$$\mathbf{x_{ij}}'\hat{\boldsymbol{\beta}} - \mathbf{x_{ij}}'\boldsymbol{\beta} + \mathbf{z_{ij}}'\hat{\hat{\mathbf{u}}}_i - \mathbf{z_{ij}}'\hat{\mathbf{u}}_i \mid \mathbf{u}_i \approx N(0, \mathbf{A}_{ij}'\boldsymbol{\Omega}\mathbf{A}_{ij}), \tag{A.7}$$

Notice here the variance-covariance matrix $\boldsymbol{\Omega}$ of the fixed-effects parameter estimates $\hat{\boldsymbol{\beta}}$ is a submatrix of $\mathcal{I}^{-1}(\boldsymbol{\psi})$.

Lemma A.3. *For* $\hat{\mathbf{u}}_i$, *we know that the asymptotic conditional behavior given* $\mathbf{u}_i$ *is the same to that in the GLMs:*

$$\hat{\mathbf{u}}_i - \mathbf{u}_i \approx \mathbf{N}(\mathbf{0}, (\mathbf{Z}_i'\mathbf{W}_0\mathbf{Z}_i)^{-1}),$$

thus, conditionally on $\mathbf{u}_i$,

$$\mathbf{z}_{ij}'\hat{\mathbf{u}}_i - \mathbf{z}_{ij}'\mathbf{u_i} \approx N(0, \mathbf{z}_{ij}'\ (\mathbf{Z}_i'\mathbf{W}_0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij}), \tag{A.8}$$

see Lemma 2.1 in Zhou and Garrido (2009a).

Note that $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $(\hat{\mathbf{u}}_{\mathbf{i}} - \mathbf{u}_{\mathbf{i}})$ are conditionally independent due to the leave one out method, we add (A.7) and (A.8) to get that conditional on $\mathbf{u}_i$:

$$\hat{\eta}_{ij} - \eta_{ij} \approx N(0, \mathbf{A}'_{ij}\boldsymbol{\Omega}\mathbf{A}_{ij} + \mathbf{z}'_{ij}\ (\mathbf{Z}'_i\mathbf{W}_0\mathbf{Z}_i)^{-1}\mathbf{z}_{ij}). \tag{A.9}$$

Appendix B

BLUE in LMM and De Vylder's Minimum Variance Estimate

In linear mixed model, the response y_{ij} is distributed as:

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i + \epsilon_{ij}, \qquad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n_i, \tag{B.1}$$

where

$$\mathbf{u}_i \ \sim \ \mathbf{N}(\mathbf{0}, \mathbf{D}), \qquad \mathbf{D} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2), \tag{B.2}$$

and

$$\epsilon_{ij} \ \sim \ N(0, \sigma_0^2/w_{ij}). \tag{B.3}$$

At the cluster level, the response is written in matrix form as:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, 2, \ldots, k, \tag{B.4}$$

where

$$\mathbf{Y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i}, \end{pmatrix}, \qquad \mathbf{X}_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{in_i} \end{pmatrix}, \qquad \mathbf{Z}_i = \begin{pmatrix} \mathbf{z}'_{i1} \\ \vdots \\ \mathbf{z}'_{in_i} \end{pmatrix}, \qquad \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, \tag{B.5}$$

and

$$\boldsymbol{\epsilon}_i \ \sim \ \mathbf{N}(0, \mathbf{V}_i), \qquad \mathbf{V}_i = diag\big(\sigma_0^2/w_{i1}, \sigma_0^2/w_{i2}, \ldots, \sigma_0^2/w_{in_i}\big). \tag{B.6}$$

On the sample level, it is written in matrix form as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \tag{B.7}$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_{k}, \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{k}, \end{pmatrix}, \qquad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_k \end{pmatrix}, \qquad \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{pmatrix},$$

$$\boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_k \end{pmatrix}, \qquad \boldsymbol{\epsilon} \sim \mathbf{N}(0, \mathbf{V}), \qquad \mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{V}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{V}_k \end{pmatrix}.$$

Lemma B.1. *Let $\boldsymbol{\epsilon}^* = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, then the linear mixed model can be expressed as:*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \qquad \boldsymbol{\epsilon}^* \sim \mathbf{N}(0, \mathbf{C}), \tag{B.8}$$

where

$$\mathbf{C} = \mathbf{Z}\mathcal{G}\mathbf{Z}' + \mathbf{V},$$

and

$$\mathcal{G} = \begin{pmatrix} \mathbf{D} & 0 & \cdots & 0 \\ 0 & \mathbf{D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D} \end{pmatrix}.$$

Lemma B.2. *The joint distribution of $\mathbf{Y}$ and $\mathbf{u}$ is multivariate normal distributed as:*

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{u} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{C} & \mathbf{Z}\mathcal{G} \\ \mathcal{G}\mathbf{Z}' & \mathcal{G} \end{pmatrix} \right). \tag{B.9}$$

Lemma B.3. *Using Lemma B.1, the MLE or weighted least square estimator of $\boldsymbol{\beta}$ is:*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{Y}. \tag{B.10}$$

$\mathbb{E}[\mathbf{u}|\mathbf{Y}] = 0 + \mathcal{G}\mathbf{Z}'\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathcal{G}\mathbf{Z}'\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ *is the best linear unbiased predictor of* $\mathbf{u}$ (**BLUP**).

Hachemeister (1975) worked on U.S data that showed linear inflation trends in claims. This trend differed from one state to the other and also from the average national inflation trend.

Frees et al. (1999) show that Hachemeister credibility regression model is one special case of linear mixed model where $\mathbf{Z}_i = \mathbf{X}_i$, for $i = 1, 2, \ldots, k$.

Lemma B.4. *Suppose that* $\mathbf{Z}_i = \mathbf{X}_i$*, for* $i = 1, 2, \ldots, k$*. Define* $\boldsymbol{\beta}_i$ *as* $\boldsymbol{\beta} + \mathbf{u}_i$*, then the generalized least square estimator of* $\boldsymbol{\beta}_i$ *based on the data in* i*th cluster is:*

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{Y}_i. \tag{B.11}$$

Lemma B.5 (Sherman-Morrison-Woodbury)**.** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are square and invertible matrices, then:*

$$(\mathbf{A} + \mathbf{XBX}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{B}^{-1} + \mathbf{X}'\,\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{-1}, \tag{B.12}$$

$$|\mathbf{A} + \mathbf{XBX}'| = |\mathbf{B}||\mathbf{A}||\mathbf{B}^{-1} + \mathbf{X}'\mathbf{A}^{-1}\mathbf{X}|. \tag{B.13}$$

Lemma B.6. *Following B.4, let* $\mathbf{C}_i = \mathbf{X}_i \, \mathbf{D} \, \mathbf{X}_i' + \mathbf{V}_i$ *denote the unconditional covariance matrix for* $\mathbf{Y}_i$*, then* $\hat{\boldsymbol{\beta}}_i$ *can be expressed as*

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i' \, \mathbf{C}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{C}_i^{-1}\mathbf{Y}_i. \tag{B.14}$$

Proof. Apply Lemma B.5 to $\mathbf{C}_i$

$$\begin{aligned}
&(\mathbf{X}_i' \, \mathbf{C}_i^{-1}\mathbf{X}_i)(\mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' \, \mathbf{V}_i^{-1} \\
=&(\mathbf{X}_i' \, (\mathbf{V}_i + \mathbf{X}_i\mathbf{D}\mathbf{X}_i')^{-1}\mathbf{X}_i)(\mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' \, \mathbf{V}_i^{-1} \\
=&\mathbf{X}_i'(\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1})\mathbf{X}_i(\mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' \, \mathbf{V}_i^{-1} \\
=&\mathbf{X}_i' \, \mathbf{V}_i^{-1} - \mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D}^{-1} + \mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' \, \mathbf{V}_i^{-1} \\
=&\mathbf{X}_i' \, (\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D}^{-1} + \mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' \, \mathbf{V}_i^{-1}) \\
=&\mathbf{X}_i' \, \mathbf{C}_i^{-1},
\end{aligned}$$

thus $(\mathbf{X}_i' \, \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1} = (\mathbf{X}_i' \, \mathbf{C}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{C}_i^{-1}$. □

Lemma B.7. *De Vylder (1981) shows that among all the linear combination of* $\hat{\boldsymbol{\beta}}_i$*, the minimum variance unbiased estimator for* $\boldsymbol{\beta}$ *is*

$$(\sum_{i=1}^{k}\mathbf{M}_i)^{-1}\sum_{i=1}^{k}\mathbf{M}_i\hat{\boldsymbol{\beta}}_i, \tag{B.15}$$

where $\mathbf{M}_i = \mathbf{D}\ \mathbf{X}_i'(\mathbf{X}_i\ \mathbf{D}\ \mathbf{X}_i' + \mathbf{V}_i)^{-1}\mathbf{X}_i = \mathbf{D}(\mathbf{D} + (\mathbf{X}_i'\ \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1})^{-1} = \mathrm{cov}(\boldsymbol{\beta}_i)\,\mathrm{cov}(\hat{\boldsymbol{\beta}}_i)^{-1}$ *is the credibility factor for* $\hat{\boldsymbol{\beta}}_i$*.*

Lemma B.8. *The MLE estimator* $\hat{\boldsymbol{\beta}}$ *in Lemma B.3 is exactly the minimum variance estimator above:*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\,\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{Y} = (\sum_{i=1}^{k}\mathbf{M}_i)^{-1}\sum_{i=1}^{k}\mathbf{M}_i\hat{\boldsymbol{\beta}}_i, \tag{B.16}$$

Proof. Use B.5, we know:

$$\begin{aligned}
&\mathbf{X}_i'(\mathbf{X}_i\ \mathbf{D}\mathbf{X}_i' + \mathbf{V}_i)^{-1}\mathbf{X}_i(\mathbf{D} + (\mathbf{X}_i'\ \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1})\\
=&\mathbf{X}_i'(\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D}^{-1} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1})\mathbf{X}_i(\mathbf{D} + (\mathbf{X}_i'\ \mathbf{V}_i^{-1}\mathbf{X}_i)^{-1})\\
=&\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{D} + \mathbf{I} - \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D}^{-1} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{D}\\
&- \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D}^{-1} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\\
=&\mathbf{I} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D} - (\mathbf{D}^{-1} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{D} + \mathbf{I}))\\
=&\mathbf{I} + \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{D} - \mathbf{D})\\
=&\mathbf{I}.
\end{aligned}$$

the rest of the proof is easy and left to the reader. □

Let $\bar{\boldsymbol{\beta}} = \frac{\sum_{i=1}^{k}\hat{\boldsymbol{\beta}}_i}{k}$. It still has the consistency and unbiasedness properties as $\hat{\boldsymbol{\beta}}$. However, this estimator is not efficient. To show $\hat{\boldsymbol{\beta}}$ is superior to $\bar{\boldsymbol{\beta}}$, we first calculate the their covariance matrices:

Lemma B.9.

1. $\mathrm{COV}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\,\mathbf{C}^{-1}\mathbf{X}) = (\sum_{i=1}^{k}\mathbf{X}_i'\mathbf{C}_i^{-1}\mathbf{X}_i)^{-1}$

2. $\mathrm{COV}(\bar{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{k}(\mathbf{X}_i'\mathbf{C}_i^{-1}\mathbf{X}_i)^{-1}}{k^2}$

In order to compare the two covariance matrix, we use the matrix form harmonic-geometric-arithmetic-mean inequality.

Lemma B.10 (harmonic-geometric-arithmetic-mean inequality)**.** *Let $w_1, \ldots, w_k$ be positive numbers such that $w_1 + \cdots + w_k = 1$, and let $\mathbf{H}_1, \ldots, \mathbf{H}_k$ be $n \times n$ positive definite Hermitian matrices. Consider weighted power means of the matrices $\mathbf{H}_i$, defined by:*

$$\mathbf{N}_s = (w_1 \mathbf{H}_1^s + \cdots + w_k \mathbf{H}_k^s)^{\frac{1}{s}}, \qquad s \neq 0, \tag{B.17}$$

and

$$\mathbf{N}_0 = \mathbf{H}_k^{1/2}(\mathbf{H}_k^{-1/2}\mathbf{H}_{k-1}^{1/2} \cdots (\mathbf{H}_3^{-1/2}\mathbf{H}_2^{1/2}(\mathbf{H}_2^{-1/2}\mathbf{H}_1\mathbf{H}_2^{-1/2})^{u_1}\mathbf{H}_2^{1/2}\mathbf{H}_3^{-1/2})^{u_2} \cdots \mathbf{H}_{k-1}^{1/2} \mathbf{H}_k^{-1/2})^{u_{k-1}}\mathbf{H}_k^{1/2}, \tag{B.18}$$

where $u_i = 1 - \dfrac{w_{i+1}}{\sum_{j=1}^{i+1} w_k}$ for $i = 1, \ldots, k-1$. Sagae and Tanabe (1994) give the inequalities:

$$\mathbf{N}_{-1} \leq \mathbf{N}_0 \leq \mathbf{N}_1 \tag{B.19}$$

Replacing $\mathbf{H}_i$ with $(\mathbf{X}_i' \, \mathbf{C}_i^{-1}\mathbf{X}_i)^{-1}$, we will find that:

$$\mathrm{COV}(\hat{\boldsymbol{\beta}}) = \frac{1}{k}\,\mathbf{N}_{-1}$$

and

$$\mathrm{COV}(\bar{\boldsymbol{\beta}}) = \frac{1}{k}\,\mathbf{N}_1.$$

According to Lemma B.10, we know that $\mathrm{COV}(\hat{\boldsymbol{\beta}}) \leq \mathrm{COV}(\bar{\boldsymbol{\beta}})$, the equality holds when $\mathbf{X}_1'\mathbf{C}_1^{-1}\mathbf{X}_1 = \mathbf{X}_2'\mathbf{C}_2^{-1}\mathbf{X}_2 = \cdots = \mathbf{X}_k'\mathbf{C}_k^{-1}\mathbf{X}_k$.

For a new subject with covariates vector $\mathbf{x}$, we can either use $\mathbf{x}'\hat{\boldsymbol{\beta}}$ or $\mathbf{x}'\bar{\boldsymbol{\beta}}$ as the estimator for its marginal mean. However the above results tell us:

$$\mathbb{V}[\mathbf{x}'\hat{\boldsymbol{\beta}}] = \mathbf{x}'\,\mathrm{COV}(\hat{\boldsymbol{\beta}})\mathbf{x} \leq \mathbf{x}'\,\mathrm{COV}(\bar{\boldsymbol{\beta}})\mathbf{x} = \mathbb{V}[\mathbf{x}'\bar{\boldsymbol{\beta}}]$$

Since $\mathbf{x}'\hat{\boldsymbol{\beta}}$ has a smaller variance, the limited fluctuation probability for $\mathbf{x}'\hat{\boldsymbol{\beta}}$ is larger than $\mathbf{x}'\bar{\boldsymbol{\beta}}$ for a given tolerance level.

Appendix C

Zero Inflated Negative Binomial

Let $\mathbf{ZINB}(y\ ;\ \pi_0, \lambda, \phi)$ denote zero inflated negative binomial distribution as in (3.9), and $\mathbf{NB}(y\ ;\ \lambda, \phi)$ denote the negative binomial distribution as in (3.10).

Lemma C.1. *Suppose y is zero inflated negative binomial distributed as $\mathbf{ZINB}(y\ ;\ \pi_0, \lambda, \phi)$, and given y, the vector $(y_0, y_1, \ldots, y_{m-1})$ follows a multinomial distributed with probability mass function:*

$$\frac{y!}{y_0!y_1! \cdots y_{m-1}!} p_0^{y_0!} \cdots p_{m-1}^{y_{m-1}!}, \tag{C.1}$$

where $y_j \in \mathbb{N}$, $y = y_0 + \ldots + y_{m-1}$ and $p_0, \ldots, p_{m-1}$ are event probabilities ($\sum_{j=0}^{m-1} p_j = 1$), then y_j is also distributed as zero inflated negative binomial distribution as $\mathbf{ZINB}(y_j\ ;\ \pi_0, \lambda p_j, \phi)$, for $j = 0, 1, \ldots, m-1$.

Lemma C.2. *Following the assumption in Lemma C.1 and suppose we only observe $y_0, \ldots, y_t$, where $t \leq m-1$, then $\sum\limits_{j=0}^{t} y_j$ is a sufficient statistics for ϕ and λ. It is distributed as $\mathbf{ZINB}(\sum\limits_{j=0}^{t} y_j\ ;\ \pi_0, \lambda \sum\limits_{j=0}^{t} p_j, \phi)$.*

Lemma C.3. *Following the assumptions in Lemma C.1 and Lemma C.2, the conditional distribution of $\sum\limits_{j=t+1}^{m-1} y_j$ given $y_0, y_1, \ldots, y_t$ is:*

$$\mathbb{P}(\sum_{j=t+1}^{m-1} y_j \mid y_0, y_1, \ldots, y_t) = \begin{cases} \mathbf{ZINB}(\sum\limits_{j=t+1}^{m-1} y_j\ ;\ \pi_{c1}, \lambda_{c1}, \phi_{c1}), & \text{if } \sum\limits_{j=0}^{t} y_j = 0, \\ \mathbf{NB}(\sum\limits_{j=t+1}^{m-1} y_j\ ;\ \lambda_{c2}, \phi_{c2}), & \text{if } \sum\limits_{j=0}^{t} y_j \neq 0, \end{cases} \tag{C.2}$$

where

$$\begin{aligned}
\pi_{c1} &= \pi_0\Big(\pi_0 + (1-\pi_0)\Big(\frac{\phi}{\phi + \lambda\sum_{j=0}^{t} p_j}\Big)^{\phi}\Big)^{-1}, \\
\lambda_{c1} &= \lambda \, \frac{\phi}{\lambda\sum\limits_{j=0}^{t} p_j + \phi} \sum_{j=t+1}^{m-1} p_j, \\
\phi_{c1} &= \phi,
\end{aligned}$$

and

$$\begin{aligned}
\lambda_{c2} &= \lambda \, \frac{\sum\limits_{j=0}^{t} y_j + \phi}{\lambda\sum\limits_{j=0}^{t} p_j + \phi} \sum_{j=t+1}^{m-1} p_j, \\
\phi_{c2} &= \phi + \sum_{j=0}^{t} y_j.
\end{aligned}$$

When we fit a zero inflated negative binomial model to the claim frequency, we use Lemma C.2 to fit the model to the truncated data, then use Lemma C.3 to estimate the future reported numbers of claims, i.e. $\hat{N}_{i,j}^{k}$.

Appendix D

Hurdle Negative Binomial Model

Let $\mathbf{HNB}(y\ ;\ \pi_0, \lambda, \phi)$ denote the hurdle negative binomial distribution as in (3.14), and $\mathbf{NB}(y\ ;\ \lambda, \phi)$ denote the negative binomial distribution as in (3.10).

Lemma D.1. *Suppose y is hurdle negative binomial distributed as $\mathbf{HNB}(y\ ;\ \pi_0, \lambda, \phi)$, and given y, the vector $(y_0, y_1, \ldots, y_{m-1})$ follows a conditional multinomial distributed with probability mass function:*

$$\frac{y!}{y_0!y_1!\cdots y_{m-1}!} p_0^{y_0!} \cdots p_{m-1}^{y_{m-1}!}, \tag{D.1}$$

where $y_j \in \mathbb{N}$, $y = y_0 + \ldots + y_{m-1}$ and $p_0, \ldots, p_{m-1}$ are event probabilities ($\sum_{j=0}^{m-1} p_j = 1$), then y_j is also distributed as the hurdle negative binomial distribution $\mathbf{HNB}(y_j\ ;\ \pi_{0j}, \lambda_{0j}, \phi)$, for $j = 0, 1, \ldots, m-1$, where

$$\begin{aligned} \pi_{0j} &= \pi_0 + (1 - \pi_0) \frac{\left(\frac{\phi}{\phi+\lambda p_j}\right)^{\phi} - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}}{1 - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}}, \\ \lambda_{0j} &= \lambda p_j. \end{aligned}$$

Lemma D.2. *Following the assumption in Lemma D.1 and suppose we only observe $y_0, \ldots, y_t$, where $t \leq m-1$, then $\sum_{j=0}^{t} y_j$ is a sufficient statistics for ϕ and λ. It is distributed as $\mathbf{HNB}(\sum_{j=0}^{t} y_j\ ;\ \pi_\Sigma, \lambda \sum_{j=0}^{t} p_j, \phi)$, where*

$$\pi_\Sigma = \pi_0 + (1 - \pi_0) \frac{\left(\frac{\phi}{\phi+\lambda \sum_{j=0}^{t} p_j}\right)^{\phi} - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}}{1 - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}}. \tag{D.2}$$

Lemma D.3. *Following the assumptions in Lemma D.1 and Lemma D.2, the conditional distribution of* $\sum\limits_{j=t+1}^{m-1} y_j$ *given* $y_0, y_1, \ldots, y_t$ *is:*

$$\mathbb{P}\Big(\sum_{j=t+1}^{m-1} y_j \mid y_0, y_1, \ldots, y_t\Big) = \begin{cases} \mathbf{HNB}\big(\sum\limits_{j=t+1}^{m-1} y_j \ ; \ \pi_{d1}, \lambda_{d1}, \phi_{d1}\big), & \textit{if } \sum\limits_{j=0}^{t} y_j = 0, \\ \mathbf{NB}\big(\sum\limits_{j=t+1}^{m-1} y_j \ ; \ \lambda_{d2}, \phi_{d2}\big), & \textit{if } \sum\limits_{j=0}^{t} y_j \neq 0, \end{cases} \tag{D.3}$$

where

$$\begin{aligned} \pi_{d1} &= \frac{\left(1 - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}\right) \pi_0}{\left(1 - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}\right) \pi_0 + \left(\left(\frac{\phi}{\phi+\lambda \sum_{j=0}^{t} p_j}\right)^{\phi} - \left(\frac{\phi}{\phi+\lambda}\right)^{\phi}\right) (1-\pi_0)}, \\ \lambda_{d1} &= \lambda \, \frac{\phi}{\lambda \sum\limits_{j=0}^{t} p_j + \phi} \sum_{j=t+1}^{m-1} p_j, \\ \phi_{d1} &= \phi, \end{aligned}$$

and

$$\begin{aligned} \lambda_{d2} &= \lambda \, \frac{\sum\limits_{j=0}^{t} y_j + \phi}{\lambda \sum\limits_{j=0}^{t} p_j + \phi} \sum_{j=t+1}^{m-1} p_j, \\ \phi_{d2} &= \phi + \sum_{j=0}^{t} y_j. \end{aligned}$$

When we fit a hurdle negative binomial model to the claim frequency, we use Lemma D.2 to fit the model to the truncated data, then use Lemma D.3 to estimate the future reported numbers of claims, i.e. $\hat{N}_{i,j}^{k}$.

Appendix E

Assumptions and Parameters in Simulations

The claim number and claim severity assumptions for different risk classes comes from Table E.1. There are 12 accident years in our simulations.

The reporting delay is assumed to be Weibull distributed, which is widely used in survival analysis. The probability of reporting delay for different risk groups are displayed in Table E.2, the maximum reporting delay is assumed to be 9.

The settlement time is assumed to be exponential distributed, unlike reporting delay, the maximum settlement delay is assumed to be 2, see Table E.3.

"Merit" and "Category" are used as regression covariates.

Risk Class	Merit	Category	Claim Count Rate	Claim Severity Rate
1	3	1	0.084	296
2	3	2	0.109	318
3	3	3	0.129	297
4	3	4	0.137	344
5	3	5	0.100	270
6	2	1	0.106	289
7	2	2	0.138	311
8	2	3	0.163	291
9	2	4	0.174	336
10	2	5	0.126	264
11	1	1	0.115	289
12	1	2	0.149	311
13	1	3	0.177	291
14	1	4	0.189	337
15	1	5	0.137	264
16	0	1	0.133	310
17	0	2	0.172	333
18	0	3	0.204	312
19	0	4	0.217	361
20	0	5	0.158	283

Table E.1: Exposure, Claim Number and Severity

Risk Class	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	0.7581	0.1422	0.0529	0.0232	0.0111	0.0056	0.0030	0.0016	0.0009	0.0005
2	0.7337	0.1498	0.0590	0.0271	0.0136	0.0072	0.0040	0.0023	0.0013	0.0008
3	0.6834	0.1622	0.0708	0.0355	0.0193	0.0110	0.0065	0.0040	0.0025	0.0016
4	0.6321	0.1709	0.0814	0.0442	0.0257	0.0157	0.0099	0.0064	0.0042	0.0029
5	0.5808	0.1756	0.0904	0.0524	0.0324	0.0209	0.0139	0.0095	0.0066	0.0047
6	0.7088	0.1564	0.0650	0.0313	0.0163	0.0090	0.0052	0.0030	0.0018	0.0011
7	0.6834	0.1622	0.0708	0.0355	0.0193	0.0110	0.0065	0.0040	0.0025	0.0016
8	0.6321	0.1709	0.0814	0.0442	0.0257	0.0157	0.0099	0.0064	0.0042	0.0029
9	0.5808	0.1756	0.0904	0.0524	0.0324	0.0209	0.0139	0.0095	0.0066	0.0047
10	0.5304	0.1767	0.0972	0.0597	0.0389	0.0264	0.0184	0.0131	0.0095	0.0070
11	0.6579	0.1670	0.0763	0.0399	0.0224	0.0133	0.0081	0.0051	0.0033	0.0022
12	0.6321	0.1709	0.0814	0.0442	0.0257	0.0157	0.0099	0.0064	0.0042	0.0029
13	0.5808	0.1756	0.0904	0.0524	0.0324	0.0209	0.0139	0.0095	0.0066	0.0047
14	0.5304	0.1767	0.0972	0.0597	0.0389	0.0264	0.0184	0.0131	0.0095	0.0070
15	0.4816	0.1745	0.1016	0.0657	0.0449	0.0318	0.0231	0.0171	0.0129	0.0098
16	0.7088	0.1564	0.0650	0.0313	0.0163	0.0090	0.0052	0.0030	0.0018	0.0011
17	0.6834	0.1622	0.0708	0.0355	0.0193	0.0110	0.0065	0.0040	0.0025	0.0016
18	0.6321	0.1709	0.0814	0.0442	0.0257	0.0157	0.0099	0.0064	0.0042	0.0029
19	0.5808	0.1756	0.0904	0.0524	0.0324	0.0209	0.0139	0.0095	0.0066	0.0047
20	0.5304	0.1767	0.0972	0.0597	0.0389	0.0264	0.0184	0.0131	0.0095	0.0070

Table E.2: Probability of Reporting Delay

Risk Class	d0	d1	d2
1	0.60	0.24	0.16
2	0.63	0.23	0.14
3	0.67	0.22	0.11
4	0.71	0.21	0.09
5	0.74	0.19	0.07
6	0.56	0.25	0.19
7	0.60	0.24	0.16
8	0.63	0.23	0.14
9	0.67	0.22	0.11
10	0.71	0.21	0.09
11	0.52	0.25	0.23
12	0.56	0.25	0.19
13	0.60	0.24	0.16
14	0.63	0.23	0.14
15	0.67	0.22	0.11
16	0.49	0.25	0.26
17	0.52	0.25	0.23
18	0.56	0.25	0.19
19	0.60	0.24	0.16
20	0.63	0.23	0.14

Table E.3: Probability of Settlement Delay

Appendix F

Simulation Study on Incurred But Not Reported Claim Number Estimation

Here we use the same assumptions and parameters as in Appendix E. Rather than focusing on the total reserve, we are mainly concerned with the estimates of the incurred but not reported claim numbers using double chain ladder and our individual chain ladder method. We give the summary table for the simulated deviations between the estimates and the empirical ones:

	Individual GLMs	DCL	SDCL
Min.	-292.00	-361.00	-367.00
1st Qu.	-47.00	-53.00	-50.00
Median	-2.00	-0.50	5.00
Mean	-0.86	-1.05	2.88
3rd Qu.	48.00	53.00	56.25
Max.	203.00	242.00	243.00
Variance	5,010.40	6,349.75	6,540.58

Table F.1: Summary of Deviations

"SDCL" represents the sum of the double chain ladder estimates based on each individual class.

We also draw the histogram, the kernel density plot and the box plot of the simulated deviations.

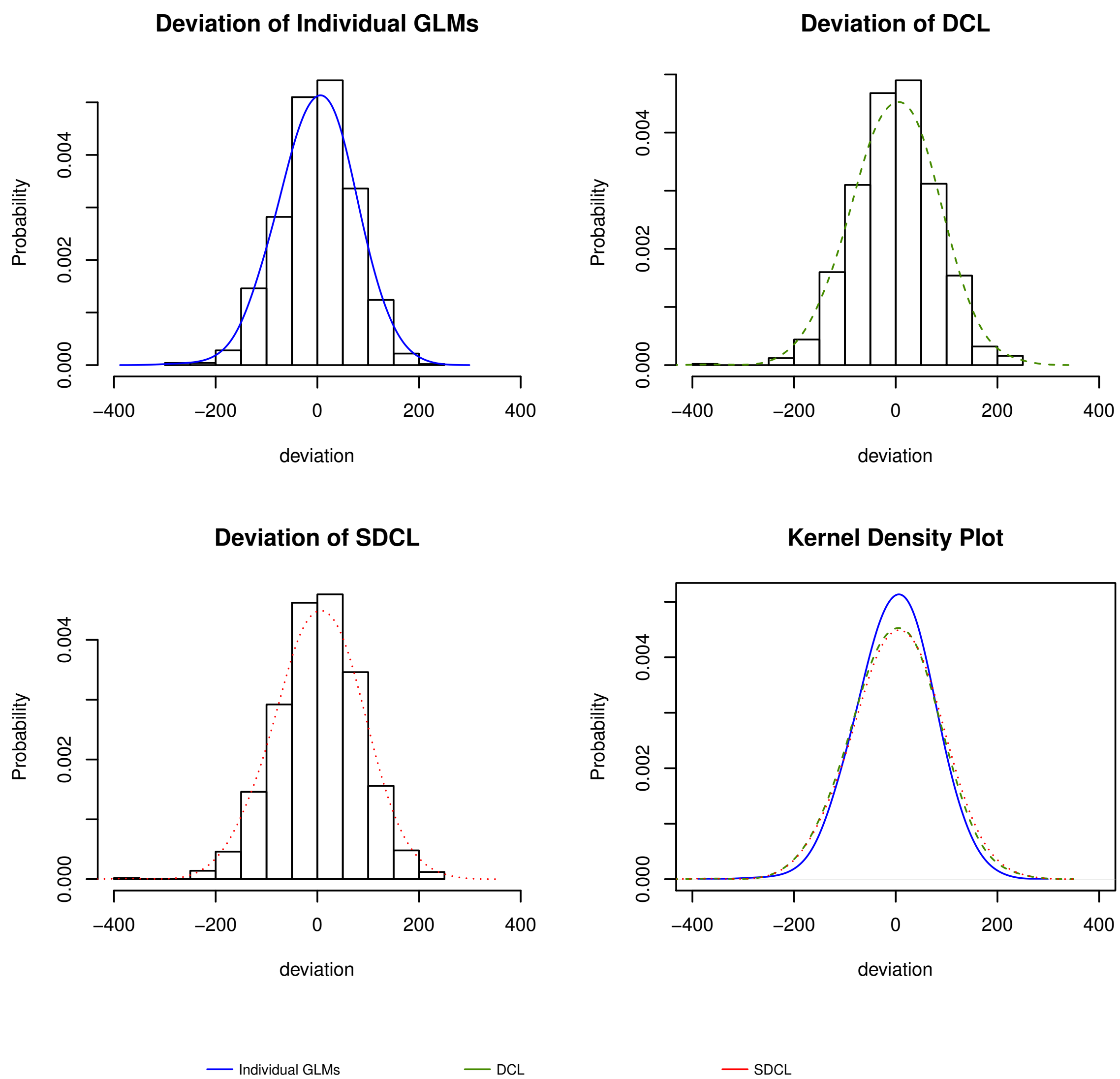

Figure F.1: Histograms and Kernel Density Plot of Deviations

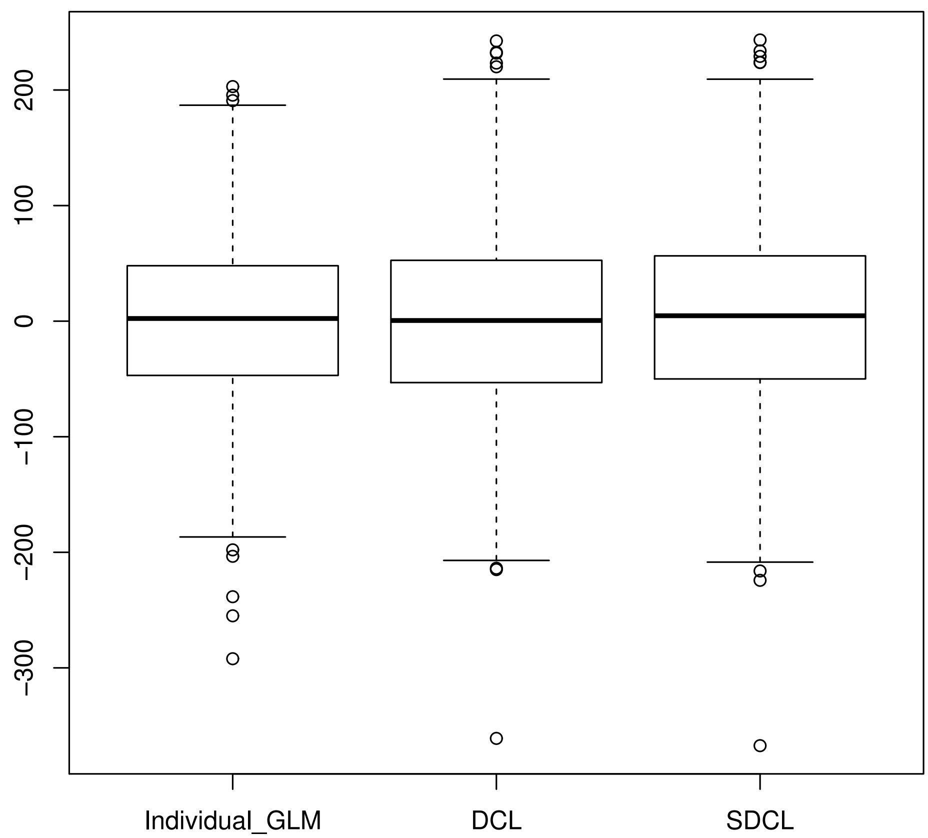

Figure F.2: Box Plot of Deviations

From the above plots, we see that DCL estimates of total incurred but not reported claim numbers have greater variance than our individual GLM estimates.

As we say in Section 3.2.4, one important reason is that the parameter α_i for the accident year effects are used in the double chain ladder method. It is over-fitted since in the assumption of our simulations, the policyholders in each class are assumed to have fixed accident rates over the accident years.